

Arms, hostages and petroleum

The inquiry into the dealings between the United States and Iran has so far concentrated on the supply of arms. What promises have been made about the price of oil?

PRESIDENT Ronald Reagan's discomfiture at the revelation in the past two weeks of his secret and possibly illicit dealings with the government of Iran is understandable. The proposition that the United States has a natural interest, but also a kind of public responsibility, to cultivate good relations with seemingly hostile governments makes sense so far as it goes, and might be regarded in the Congress as justification for an attempt at rapprochement with Vietnam, North Korea or (come to that) the Soviet Union. But Iran? With the memory of the capture of the staff of the Tehran embassy's staff still so fresh? That alone would expose Mr Reagan to the complaint that deals with hostage-takers can serve only to put further potential hostages at risk. So much is so clear that it is inevitable that people should believe what the president denies, that the "small shipment" of arms to Iran was a consideration for Iranian influence with the captors of several US hostages still held in the anarchic Lebanon. No doubt the tail-enders of the 99th Congress will get nearer to knowing what really happened.

Meanwhile, there is a larger question to be explored than that of whether some part of the US administration was prepared to make a deal most other people would have considered unwise. When the news filtered from Iran at the beginning of this month that Mr Robert McFarlane, director of the US National Security Council until a year ago, had taken to Tehran not merely a small shipment of arms, but a bible and a cake, the word was that the United States was also willing to use its influence to bring about a general increase in the price of oil, from something like \$15 a barrel now to something like \$18. While attention centres on the arms and the bible (it is harder to imagine what symbolic purpose the cake may have served), the price of oil tends to be overlooked, yet on the assumption that dealing over hostages will now become unfashionable, this is a more lasting and more serious issue.

Circumstantial evidence suggests that something odd may have been going on. As a member of OPEC (the Organization of Petroleum Exporting Countries), Iran has recently vociferously expressed its discontent at the fall by more than a half, even in nominal terms, of the price of oil on the international market. Its own income from petroleum has been cut not merely on this account, but by interruptions of the movements of oil tankers from its terminals at the head of the Persian Gulf. But for the past several years, as for most of the time since OPEC's beginning in 1969, the market in oil has been determined by the policies of Saudi Arabia, whose reserves of cheap oil are now even more dominant than before 30 years of vigorous exploitation in the Middle East. Apparently exasperated by OPEC's failure to agree on production quotas tight enough to sustain a higher price, Saudi Arabia has for the past two years gone for market share instead, increasing its own production and letting the price find its own level to the general benefit of the industrialized West and the discomfort of relatively high-cost producers of oil, Iran for one and Britain (not a member of OPEC) for another.

That policy has now been turned on its head. Sheikh Ahmad Zaki Yamani was dismissed his post as oil minister of Saudi Arabia a month ago, since when his successor has arranged a

meeting of OPEC advertised as one at which the price of oil should be increased. There are several explanations, of which one is that the United States has successfully exerted its influence with the Saudis, in the manner suggested by the shadowy (but at least partly accurate) sources in Tehran. It goes without saying that this sequence of events affecting the price of oil may be as plausibly explained by Saudi exasperation that the United States had begun dealing with Iran again; Saudi Arabia may even have resolved that, if anyone deals with Iran, it should do so itself. There are also many other explanations in which President Reagan's little shipment of arms plays no part.

For the time being, the US administration's interest is that it should if possible be cleared of the suspicion that it has been prepared to monkey with the price of oil on the international markets. In its way, this charge is as serious as is the charge that the United States has been prepared to trade with those who support or encourage kidnappers, thereby helping to institutionalize political hostage-taking. Not that Iran would be the only beneficiary of even a modest increase of the price of oil. The British chancellor of the exchequer would benefit, as would his opposite numbers in Nigeria and Mexico (not to mention those who worry about Mexico's debts). So too would US oil producers in Texas and California. But the great majority of those who buy OPEC oil, from Japan to Western Europe, would be hurt by an increased price of oil, while the flickers of hope that economic growth can yet be rekindled would become more fitful. These are the players who will want to know what, if anything, President Reagan and his emissaries have been telling the oil producers of the Middle East. □

Insiders outlawed

Cleansing the world's financial markets of fraud may be more difficult than supposed.

THE wave of righteous indignation about insider trading on Wall Street and, on a lesser scale, in London, is understandable but suspect. Not that those who use privileged information to play the stock markets to their own advantage are distinguishable from common thieves. Mr Ivan F. Boesky, the New York financier required by the US Securities and Exchange Commission two weeks ago to hand over \$50 million in ill-gotten gains and to pay the same amount as a penalty, is the most spectacular case so far. But the case of Mr Geoffrey Collier, the director of Morgan Grenfell in London who was forced to resign his job after seeking to make £15,000 on a single insider trade, is more illuminating if only because it is easier to understand.

Collier's fault was to have sought to buy shares in a British engineering company in advance of a takeover bid of which he had private knowledge, gathered in the course of his work. The deal was executed on his behalf by an overseas company and through a New York stockbroker, which accepted the order to supply the stock at an agreed price even though it did not have enough on its books at the time. The stockbroker's calculation was that it could have bought the stock in the markets later in the day, but in the event it laid off two-thirds of the risk that the price



might rise onto another New York firm, a subsidiary of Chase Manhattan. If the deal had not been cancelled, one firm would have lost £5,000 and the other £10,000.

So what? will be one cynical question. Stockbrokers are forever making huge sums of money by "dealing their books", so why weep when they risk a loss? The simple answer is that it is not their money that they risk, but ours, that invested in the international stock markets by pensions funds, savings institutions and by banks that take deposits from the general public. If Collier had pocketed his £15,000, the ultimate losers would have been ordinary people. Boesky's \$50 million ultimately derives from the same source. But \$50 million is only a tiny fraction of the funds belonging to the public invested in the international financial markets. Is the rest safe?

What worries the Securities and Exchange Commission now is that, in the United States, the huge amount of paper debt obligations called "junk bonds", perhaps as much as \$100,000 million, which has sustained the recent wave of spectacular takeovers, may be suspect. These securities (if that is the right word) entitle the owner to high interest, but are not secured against assets and not traded on stock exchanges. Holders can get back their capital by appealing to the investment banks that issued the junk bonds, but the fear is that this procedure may have been infected by Boesky get-rich-easy principles; that could spell disaster, given the extent to which savings and loan associations in the United States have traded their clients' cash for junk bonds (which look better on a balance sheet).

So there is a strong case for making outlaws of the insider-traders. How is that to be done? This is the point at which righteous indignation comes a cropper. The case of Collier is straightforward; he acquired his inside knowledge because the firm of which he was a member had agreed to mount a bid for the British engineering company (on behalf of Mr Robert Maxwell, the proprietor of Pergamon Press). To seek to profit personally from this knowledge was a violation of his firm's internal regulations, of the London Stock Exchange's rules and of British government law (hurriedly brought into effect in time to put Collier under pressure). Boesky's is a more complicated case; so far, he has been accused only of using inside information provided by a true insider, Mr Denis Levine, arrested by the Securities and Exchange Commission earlier this year and who has since traded immunity from further prosecution for the repayment of \$1.5 million and continued cooperation with the commission's inquiries; who, in law, is the real felon, Boesky or Levine?

The truth is even more complicated. Without accepting the arguments of the Chicago economists who say that insider trading is virtuous because it brings reality more speedily into effect and thus make the markets more efficient, it is fair to ask whether stockbrokers who commission studies of industrial sectors, ostensibly for the benefit of clients, should then be allowed to speculate in the stocks concerned. And should the 'analysts' who carry out the research be allowed to speculate on their own accounts, knowing as they must that at least some of their company's clients will believe what they say? When is it improper to act on (and profit from) rumour? Or even to take advantage of special knowledge of the health (or sickness) of publicly-quoted companies gathered in quite different connections, as a journalist, for example?

The righteously indignant will quickly say that there should be no limits to the most simplistic writ. (Simplistic Senator William Proxmire is alarmed at the way in which corporations have been saddled with huge amounts of debt after successfully defending against a takeover bid, but that is another issue for which directors are responsible.) Others will recognize that there is an urgent need for a legal definition of a boundary between right and wrong that will make sense, and will endure. Governments, poor things, will know that the problem is complicated by the international character that the financial markets have now acquired. □

Teachers in travail

Is the proper education of the young too important to be left to teachers?

LIKE motherhood in the old days, the education of the young has become an important social objective, and for good reason; economic survival depends on social or technical innovation, which requires innovative young people. So it is not surprising that governments in the industrialized West should be seeking to improve high-school education. In the United States, the new Democratic Congress has lost little time in saying that it wants the federal government to do more for state-supported school systems. (What will happen to the Gramm-Rudman budget deficit reduction act if the defence budget is not cut?) In Britain, the government is in the last bruising stages of an attempt to trade money (which teachers need) for a promise (which teachers' labour unions are reluctant to concede) that more able teachers will be better rewarded. Everywhere, the simple arithmetic is overlooked.

Here is the arithmetic. Most Western countries send their children to school for 12 or 13 years, in classes that average about 20 students per teacher. In round numbers, the schooling of each child requires roughly two-thirds of a teacher-year. But teachers' professional lives last for about 35, and are statistically much shorter, which means that for a demographically unchanging population, at least 2 per cent of the population of working age should be schoolteachers. Naturally, governments and the societies they represent require that teachers should have had the benefit of higher education, in which the participation rate in Britain (admittedly one of the worst) is merely 15 per cent. If all those leaving higher education followed professional careers for 35 years (which is far from being true), one seventh of the population would have to be schoolteachers. Allowing for the manpower needs of higher education, one should expect that roughly 20 per cent of college-leavers would be occupied in education. Is it reasonable to expect that those who teach the young will also be among the most talented?

There are three distinct answers to this question. In the United States, where the federal government has, for the past six years, stuck largely to exhortation, but where the quality of high-school education has become a general anxiety, some state governments are making valiant efforts to improve the performance of their schools, while institutions that train teachers are trying to enhance the reputations of those who teach in schools by requiring more from entrants to their courses. These private initiatives have not so far been well-rewarded because people will not pay particularly well-qualified teachers extra, while teachers themselves pretend that discrimination is anathema. In Britain, on the other hand, the government is willing to battle for its conception of a core curriculum and is also pleading for pay-discrimination; it should be known this week or next whether its offer of an extra 17 per cent over 18 months will be accepted on the teachers' terms or will have to be enforced. Only in Japan is there a workable if unstable solution of the problem: put more than a third of the population through higher education, but then arrange that married women do not work, but stay home and teach their children.

Elsewhere, governments will have to live with solutions far from their ideals. In North America and Western Europe, the chance of persuading the best and brightest of those leaving higher education to teach in schools is small, especially in those fields from which technology springs. Better to be reconciled to that than to keep tilting at a windmill. So the crying need is for a general recognition that schooling is not an end in itself but one of several means to functional adulthood, that the core curriculum should be stripped down so that it is no longer a test of the preternatural ability of gifted students but within the competence of all students and most teachers and that, whatever governments do, the results will not be perfect. □

Europe's framework research

Commission ready to pull out communal rug?

Brussels

THE European Commission could go so far as to drop the whole \$8,000 million, 'framework five-year programme for European research and development' next month, if governments fail to agree to support the programme at a sufficiently high level. So observers were saying in Brussels last week, as the Commission began to field its strongest political musclemen, Commission President Jacques Delors and his team, in favour of what the Commission considers a sensible budget for the research programme.

Last week, the European Parliament's Energy, Research and Technology Committee voted to advise the Commission to withdraw the framework programme if the full budget requested — a reduction compared with the \$10,000 million earlier requested and a "minimum" according to the committee — was not accepted.

Certainly research, which at the full amount of the framework programme would account for little more than a twentieth of Brussels spending, is now at the centre of Brussels politics, where it is feeling the full blasts of European summitry. The Commission wishes through research to confirm a new role for itself in the industrial regeneration of Europe. France, West Germany and the United Kingdom have been resisting both for financial reasons and, undoubtedly, because of the interference in national science and technology programmes that increased Commission activity would imply.

In fact, governments have accepted the principle of a major European program-

me of research and development coordinated by the Commission in the 'Single Act', the major treaty of union agreed by heads of state last year in Luxembourg and due to be voted in place by all members of the European Communities by the end of this year. The Single Act includes a place for the framework programme for research, and Delors sees the voting and proper financing of that programme to be a clear test of the 'New Europe'. He told British industrialists recently that European Communities research and development policy was "so relevant and so essential" because in the long term it bore directly on the creation of a single European market for European goods.

The framework amounts to less than 2 per cent of overall research and development in Community countries and "we are not, repeat not, out to enter into some kind of competition with the member states", Delors said. "We hear references to cost-effectiveness. But the microchip industry people tell us that their return on ESPRIT programmes is 400 per cent despite the cost of cooperation. We are said to be bureaucratic. But do people know that it takes two European officials to manage £1 million of research and development funds, and eight officials in the most efficient member states? Industrialists, and researchers, and academics are turning to us, to Europe, and saying they are ready to work together... I do not see why the most solemn undertakings by the heads of state and government [in the Single Act] are not being put into effect faster."

Robert Walgate

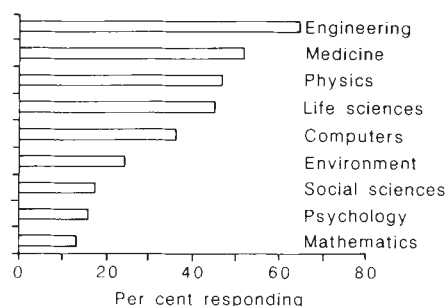
US universities appear in good shape

Washington

CONSTRUCTION in progress and planned over the next five years will increase research space at US research universities by between 14 and 26 per cent, according to a survey* by the National Science Foundation, at a cost of \$7,500 million.

The survey covered all doctorate-granting institutions, and showed that the top 50 in terms of research and development spending (out of a total of 165) accounted for more than 50 per cent of work in progress and 60 per cent of planned work. Engineering, medical sciences and physical sciences departments were more likely to have construction in progress than others (see figure).

The survey produced one surprise for the doomsayers on university research



facilities: over half of the same top 50 universities reported the condition of their research facilities as "good" or "excellent".

But of those ranked outside the top 50, over half replied "fair" or "poor". □

*Science and Engineering Research Facilities at Doctorate-granting Institutions. National Science Foundation, 1986.

UK research appointments

Council breaks new ground

THE terms on which the Natural Environment Research Council (NERC) has made three senior appointments to its staff have occasioned some surprise among British academics.

Interest centres particularly on the arrangements under which the three directors of science appointed in the past twelve months have been promised that, if their five-year terms are not renewed, they will be kept on the payroll at a professional level.

The three appointments were foreseen in NERC's corporate plan, published nearly two years ago. The objective was that NERC's spending in three areas (Earth science, marine science and terrestrial and freshwater ecology) spanning its whole range of interests should be overseen by scientists of individual distinction. When the new arrangements were announced, some directors of NERC institutes expressed misgivings on the ground that their independence would be compromised and their access to the NERC council impeded.

The first appointment, of Dr P.B.H. Tinker (terrestrial and freshwater ecology) was made last December. Dr J.D. Woods (marine science), previously at the Kiel institute of oceanography, and Dr J.C. Briden, previously professor of geology at the University of Leeds, were appointed in the summer of this year.

As part of the arrangement with the holders of the three posts, it has been agreed that they should remain active in research, for which reason it is intended that they should maintain modest research programmes at suitable academic institutions (All three have plumped for the University of Oxford, within driving distance of NERC's Swindon headquarters.) Dr J.C. Bowman, secretary of NERC, says on the telephone that the council of NERC has agreed that sums "of the order of 100K [£100,000]", taken from the budgets of the three directorates, should be used to support this research.

On NERC's promise of continued employment, Bowman says that the arrangements do not differ from those governing the appointment of scientists elsewhere in the council's network, who are normally offered employment until retirement age. Moreover, he says that the arrangements are not inconsistent with the British Treasury's rules on public appointments, but merely a "new wrinkle" on them.

Tinker will be 60, the normal retirement age from the public service, when his five-year appointment ends, but both Woods and Briden (now 47) will have some years to go.

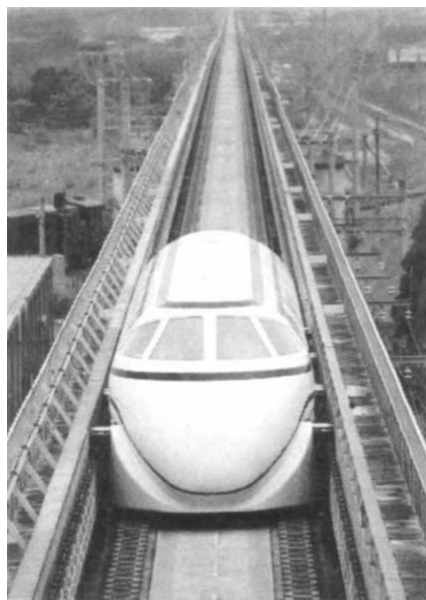
John Maddox

Magnetic levitation

Japan's railways look ahead

Tokyo

HIGH-SPEED magnetically levitated (Maglev) trains are in the news again. When Japan, which has been running a modest Maglev research programme, invented a



517 km per hour experimental train in 1979, few thought it could ever emerge from the test track. But the climate has suddenly begun to change and politicians are openly talking of a magnetically levitated bullet train that could transport passengers from Tokyo to Osaka in little more than an hour.

A prototype train that can carry up to 80 passengers is now under construction. Wind tunnel tests on a 7-m scale model of the prototype were successfully carried out at Nissan Motor Company's Technical Centre two weeks ago, and the prototype itself should be ready for test runs on Japanese National Railways (JNR)'s 7-km test track in Mihazaki, Kyushu, next spring.

The break-up and privatization of JNR

has been the catalyst. Years of profligate government spending and intransigent unions have left JNR with debts of \$230,000 million, more than the combined national debts of Mexico and Brazil. Interest payments alone have been running at more than \$20 million a day.

Eight bills have raised through the Lower House of the Diet tailored to break up JNR into six regional passenger railways, a freight railway and a holding organization for the profitable bullet train lines (Shinkansen) which will be leased back to the passenger railways. More than 3,000 hectares of JNR land, much of it in prime areas, will be sold off to recoup some of the debt and 61,000 JNR employees will have to find employment elsewhere. Yet apart from a few radical attacks on the railways communication system which brought temporary chaos to commuters (see *Nature* 318, 405; 1985), opposition to the privatization plan has failed to materialize and many union members have capitulated for fear of losing their jobs.

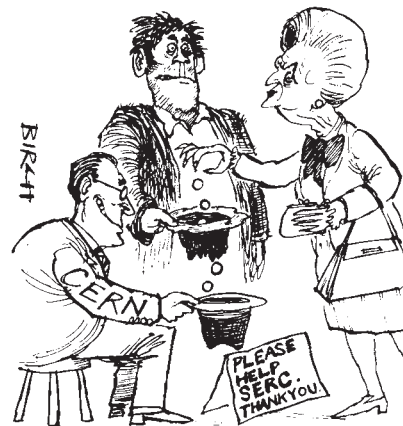
After the break-up in April 1987, JNR's research laboratories will be reborn as a research institute backed by a foundation. More money is to go into Maglev. Maps for new Shinkansen lines, pencilled in during Japan's high-growth era, are being taken out of the back drawers of filing cabinets for reappraisal. It is hoped that one of these lines could carry a Maglev train.

The Maglev train floats on the magnetic field of superconducting niobium-titanium magnets mounted below the car floor. As the magnets pass over aluminium coils set in the U-shaped trackway, electric currents induced in the coils set up a repulsive magnetic field off the same polarity that levitates the vehicle about 10 cm off the ground. Guidance and propulsion are provided by electromagnets set in the walls of the trackway.

The train offers a smooth, silent and

SERC scoops pool

EARLIER this week, it seemed certain that the Science and Engineering Research Council (SERC) would be given the lion's share of the extra £25 million for research provided three weeks ago by the British government. This appears to have been the



decision of a meeting last week of the independent members of the Advisory Board for the Research Councils (ABRC), the government's cake-cutting committee, and is unlikely to be overturned by a meeting of the full committee arranged for Wednesday this week.

SERC's success is a pyrrhic victory. The extra funds channelled in its direction will be used exclusively to pay the increased cost (in sterling) of subscriptions to overseas research organizations, among which the European high-energy physics research centre at Geneva (CERN) is the largest. The cost of the CERN subscription, now roughly £45 million, has risen because of the devaluation of sterling relative to the Swiss franc. □

bump-free ride at more than 400 km per hour, and in commercial service will be controlled by ground stations at 50-km intervals along the track. One slight problem, however, remains. In an earlier test, the maximum magnetic flux on the floor of the car was 200 gauss. Passengers complained that their watches stopped. In the prototype the disposition of the train's superconducting magnets will be adjusted to try to overcome this effect, according to Yoshihiro Kyotani, manager of the Maglev project.

Kyotani already has the most likely route for the first commercial Maglev line dotted on a map in his office, the Chuo (Central) Shinkansen, which would run from Tokyo to Osaka via Nagoya.

Although nobody is yet discussing construction costs, Kyotani points out that the trackway for the new vehicle would cost less than that for a conventional railway. But he admits that the cost of the Maglev vehicle itself will probably be higher than that of an ordinary train.

David Swinbanks

Soviet reservations on IAEA conventions

THE Supreme Soviet has approved the two new conventions of the International Atomic Energy Agency (IAEA) formulated during the summer in the light of the Chernobyl catastrophe. The signing is hardly remarkable; the effect of Chernobyl was to crystallize ideas that had begun to take shape among IAEA members after the accident at Three Mile Island. But Soviet media and government spokesmen have consistently claimed the new conventions as a specifically Soviet initiative.

But the Soviet Union has ratified the treaties only in part. A special reservation has been added to the effect that the USSR

will not consider itself bound by the provisions of Article 11.2 of the Convention on Prompt Notification of a Nuclear Accident nor by Article 13 of the Convention on Assistance in the Event of a Nuclear Accident or Radiation Emergency. These provisions refer to the possibility of submitting any dispute arising under the terms of the conventions to arbitration or to an international court at the request of any side. The Soviet Union, says the reservation, declares that the consent of all sides is necessary before any dispute is referred to arbitration or to an international court.

Vera Rich

US polar research

NSF break with Coastguard?

Washington

PLANS for polar research have been upset by the US Coastguard's recent downgrading of one of its ships, *Glacier*, which is no longer classified as an "icebreaker", but described as "ice-strengthened" instead. As a result, the National Science Foundation (NSF) may break with the US Coastguard, and find polar research vessels elsewhere.

Glacier was considered the most suitable of the Coastguard's icebreakers for research, and the discovery that its hull has been seriously weakened by corrosion has re-opened the debate about whether Coastguard ships are still suitable as research platforms for NSF. Two new Coastguard icebreakers should be completed in 1992 or 1993, but NSF will have to span the gap from now until then.

The nub of the problem is that Coastguard ships are primarily military vessels not well equipped for research. Two years ago, NSF leased one small research icebreaker, *Polar Duke*, and is also now taking research space on the West German icebreaker *Polarstern*, which has superior research facilities. But restoring *Glacier* to icebreaker status would cost

\$20 million and NSF considers the cost would be too great.

Instead, officials want the use of a dedicated research icebreaker from another source, either by collaboration with another country or by leasing. Without NSF support, the Coastguard cannot justify large expenditures on *Glacier*, although it says no final decision has been taken. Loss of the NSF account would mean about \$2.3 million a year.

Scientific interest in Antarctica seems to be increasing year by year. Investigation of the Antarctic ozone hole seems now to guarantee that interest will not wane. NSF, as well as using *Glacier*, has also frequently used two other Coastguard icebreakers, *Polar Star* and *Polar Sea*. But these are unpopular with researchers and in need of refits.

Glacier is working this season in the Antarctic ice edge zone, pushing but not breaking ice. Research has not been unduly affected this year, but it could be if NSF were tied to the ship. The two new vessels due next decade will provide better accommodation and facilities than any existing Coastguard ship. Meanwhile, NSF is looking at its options. **Tim Beardsley**

First leptonic collisions at Japan's KEK

Tokyo

TRISTAN, the giant accelerator at Japan's High Energy Physics Laboratory (KEK), burst into action last week when electrons and positrons were smashed together at a record-breaking centre-of-mass energy of 50 GeV in the accelerator's 3-km main storage ring. And with foreign competition breathing down its neck, the TRISTAN team is now racing to get its experimental programme into full swing.

After successful tests of the electron and positron beams at 25.5 GeV earlier this month (see *Nature* 324, 7; 1986), KEK scientists set the counter-rotating beams in the main ring on collision course and the first hadronic-like event was picked up by the partially instrumented VENUS detector shortly before midnight on 18 November, followed a few hours later by a collision producing large-angle Bhabha scattering.

The record set by TRISTAN exceeds the maximum of 47 GeV achieved by the PETRA ring in Hamburg. But the maximum luminosity obtained, $2 \times 10^{29} \text{ cm}^{-2} \text{ s}^{-1}$, is barely a hundredth of TRISTAN's design capability, and the electron- and positron-beam currents, at present operating at around 1 mA, will have to be raised before a serious search for new elementary particles can begin.

AMY, a detector built in collaboration



Aerial view of the TRISTAN complex.

with teams from the United States, Korea, China and Japan, has just been rolled into position in one of the four experimental halls on the main ring. And the TOPAZ detector, now undergoing tests with cosmic rays, should be in place next spring ready for the full experimental programme to begin in early May 1987.

TRISTAN thus has a head start on its most immediate rival, the Stanford Linear Collider, whose even higher energy electron-positron collisions are not expected for at least a few months. Even then, the linear collider will be limited to a single energy, producing showers of intermediate vector bosons (the Z^0), TRISTAN will have about two years as the world's most powerful flexible electron-positron ring collider until the LEP ring at CERN comes on line in 1988. **David Swinbanks**

Optical interferometry

Large grant for new instrument

Sydney

THE University of Sydney has won a grant of A\$500,000 to begin construction of a 640-m baseline high-resolution stellar interferometer. This development is a milestone for stellar astronomy (and for the Chatterton department of astronomy at Sydney) as well as for the Australian Research Grants Scheme, whose budget is so small that projects on this scale have so far been beyond its scope.

The new instrument will be a larger version of a prototype instrument built by Sydney over the past 15 years. Only about 30 stars, mostly bright giants, have so far had their apparent diameters measured to an accuracy within 5 per cent by this technique. The new interferometer telescope will be able to measure the apparent diameter of 50,000 stars (brighter than 8th magnitude) with an accuracy of 2 per cent or better.

Measurement of the apparent size of stars whose distance is known should allow the establishment of fundamental scales of stellar diameters and luminosities to be established. Observations of close binary stars can give all the parameters of the system, including their orbits, temperatures, luminosities, diameters and distances. Other problems the interferometer could help solve include mass loss, by observing how a star's envelope appears to grow at longer wavelengths, and stellar rotation, by observing the flattening of the poles.

All the essential technology for the interferometer telescope has been proved in a prototype developed by the Sydney University team, led by Dr John Davis, over the past 10 years. The prototype has an 11.4-m baseline with a steerable mirror at either end. The mirrors capture the light from a star which is then reflected to a central optical bench where the beams are combined and interference fringes observed. The prototype will form the heart of the new instrument.

The principle on which the new interferometer will operate is amplitude interference, its design being a development of Michelson's classic interferometer. Amplitude interferometry has not previously been a workable option because of the blurring caused by atmospheric 'seeing' effects. Until this year the Australian Research Grants Scheme has been restricted to grants of less than A\$150,000. The \$150,000 grant for the interferometer marks the beginning of the government's policy of more funds for the scheme concentrated into fewer but more highly rewarding grants (see *Nature* 323, 98; 1986). **Charles Morgan**

Environmental databases

NOAA slipping further behind

Washington

THE attempts by the US National Oceanic and Atmospheric Administration (NOAA) to bring up to date its over-stretched environmental data archiving and distribution system have run into trouble. The agency's proposed new system, NOAAANET, is intended to supply retrospective environmental data from satellite and other sensors for researchers during the 1990s and into next century. But political obstacles have prevented NOAAANET from finding its way into the agency's budget for fiscal year 1988, which starts next October, and there are fears that the administration's outmoded technology could slip even further behind.

NOAAANET was cut out of the agency's 1987 budget by the Office of Management and Budget. This year, it did not even get that far: NOAA has been told by its parent agency, the Department of Commerce, to see whether the system could not be run by the private sector. Nonplussed NOAA officials say prospects for making a profit out of environmental data archives are not immediately clear.

Virtually all researchers agree that NOAA has fallen way behind in the push to computerize access to environmental data, even though it has a legal mandate to provide such an archive. Requests for data can take months to be processed, and there is no on-line catalogue, let alone data access. The problem will worsen as new advanced satellites will generate data in ever-increasing quantities. The National Aeronautics and Space Administration (NASA), in contrast, although supposedly a research agency, has already established data centres with much more advanced capabilities and is starting to link them up. The new 'NASAnet' will enable researchers to examine catalogues of data from different sensors across various disciplines and make selections from their computer terminals.

The difficulties result, some say, from NOAA's disadvantaged position within the Department of Commerce. The department's bureaucrats are less likely to be sympathetic to projects benefiting research than to those of NASA, for example. NASA's initiatives to make use of computers for data handling and distribution have been generously funded: the agency recently signed a \$240 million contract with Boeing Computer Services for project support for a new network.

There are nevertheless those who are critical of NOAA's preliminary plans for NOAAANET. Frank Eden of the Joint Oceanographic Institution questions whether there have been adequate consultations with NASA and the National Science Foundation, for example, which is

also constructing a large computer network based around its five supercomputer centres.

The lack of a formal scientific advisory body to advise NOAA on how to implement the project is also a source of concern. But Gregory W. Hunolt, NOAAANET project manager, says NOAA plans to bring in external advice

shortly; meetings with NASA have already started and a definition document has been circulated in the academic community. High-level contacts between the agencies had been delayed by a NOAA proposal to put its orbiting meteorological instruments on a private satellite, Omnistar, but that idea has now been dropped. But without a commitment by the Department of Commerce to seek the several tens of millions of dollars that NOAAANET would cost, progress seems unlikely.

Tim Beardsley

AIDS

WHO plans megaprogramme

Washington

THE World Health Organization (WHO) last week said that prevention and control of AIDS (acquired immune deficiency syndrome) has become one of its top priorities and that efforts to combat the disease need to be as urgent as those against smallpox in the 1950s. WHO estimates there will be between 500,000 and 3 million AIDS cases by 1991, with up to 100 million people infected with the virus.

Dr Jonathan Mann, head of WHO's AIDS programme, says those figures could be serious underestimates if AIDS spreads quickly in Asia and in South

America. WHO estimates the worldwide cost of fighting the disease at \$12,500 million by the early 1990s.

The WHO AIDS programme officially started in May of this year, and has already raised \$4.5 million in special contributions from member governments. The agency is seeking to raise \$200 million for next year. By next decade it wants \$1,500 million a year, 50 per cent more than its entire budget at present.

WHO will concentrate on the exchange of information among researchers working on prevention and therapy, as well on direct assistance to countries needing to set up their own control programmes. The research component will include coordination of vaccine field trials and the establishment of an international virus bank.

Mann estimates that about 100 countries will need technical assistance in setting up their own prevention and control programmes. These should be based on research on how the disease is transmitted and public education, and should include "hard-headed" social science research into the availability and usage of condoms and hypodermic needles, for example. WHO will also foster operational research into the development of cheap antibody testing for blood banks worldwide.

Tim Beardsley

- The British government last week launched a £20 million campaign to educate the public on the danger of AIDS. "Don't die of ignorance" is one of the slogans that are to confront the British public in a wave of publicity including television, radio and maildrop advertising. The campaign began this week with national newspaper advertisements.

At the end of the year, a month-long television campaign using a 40-second commercial will pave the way for the maildrop. In the new year the 23 million homes in the United Kingdom each receive a letter containing an explicit warning on the danger of AIDS.

A minute-long radio commercial will also be broadcast at the end of the year, narrated by a top disk jockey to attract the attention of the young.

Bill Johnstone

India screens foreign students for AIDS

New Delhi

MEDICAL screening for AIDS has been made compulsory for all foreign students seeking admission to Indian educational institutions. Under the new guidelines from the health ministry, foreign students provisionally admitted may be liable for deportation if tests confirm the presence of antibodies to human immune deficiency virus (HIV).

The government directive was issued last month after the detection of HIV antibodies in an African student in Madras. He is the only male seropositive case among 22 cases so far identified in the city.

India has an estimated 30,000 foreign students, mostly from Africa and eastern Asia. AIDS screening has been proposed for fresh admissions, but is likely to be made mandatory for foreign students already here. The Indian Council of Medical Research has set up 19 centres to screen sera using imported ELISA (enzyme-linked immunosorbent assay) kits.

There have so far been three deaths from AIDS in India — two of the patients had received blood transfusions in the United States and one was an Indian settled in Malawi. Some 30 antibody-positive cases, most of them prostitutes, are under observation.

K. S. Jayaraman

Japanese volcano

Islanders flee surprise eruption

Tokyo

JAPAN's volcanologists were taken by surprise when Mount Mihara on Izu-Oshima island south of Tokyo exploded in a series of violent eruptions last Friday, forcing the overnight evacuation of the island's 10,000 inhabitants as a river of lava threatened to engulf the main town.

Izu-Oshima is the first in a chain of volcanic islands that extends southward for more than 1,000 km into the Pacific along the edge of the Philippine plate. A submarine volcano at the southernmost end of the volcanic arc temporarily broke surface during an eruption earlier this year (*Nature* 319, 441; 1986), and Miyakejima, Izu Oshima's immediate neighbour, erupted in 1982. But Izu-Oshima itself has been silent for the past 12 years, and an eruption on the present scale has not occurred in the past two centuries.

The first signs of activity came in July when the island was rocked by volcanic tremors. But surveys by the Meteorological Agency and Tokyo University's Earthquake Research Institute showed no evidence of the 'bulging' of the Earth's crust

residents were in no immediate danger. Hours later the evacuation order came.

What nobody had foreseen was the opening of a new line of vents that cut through the outer rim, something that had not occurred in more than 500 years. As the new vents burst into life, lighting up the night sky in a curtain of fire, a river of lava set forests ablaze as it surged towards the main town of Motomachi.

By dawn the island was deserted, the 10,000 residents and 2,000 tourists having been evacuated to the mainland. All that remained were abandoned domestic animals, a farmer who stubbornly refused to leave his 5,000 chickens, and some baffled volcano experts.

On Monday, the lava front lay within a few hundred metres of Motomachi, but volcanic activity seemed to have switched to the southeast coast of the island, where plumes of discoloured water in the sea suggest submarine volcanic activity.

David Swinbanks

Biotechnology insured

Response to product liability

Washington

THE high cost of US product liability insurance has stimulated the Association of Biotechnology Companies (ABC) to plan a captive cooperative insurance company for the industry. If enough companies join the risk-sharing enterprise, it could offer "attractive" rates for insurance in respect of product liability as well as for the actions of directors and officers of companies, according to Bruce Mackler, ABC's general counsel.

Many smaller biotechnology companies have apparently not been able to obtain insurance at reasonable cost; some have had to proceed with none. ABC says one (unnamed) biotechnology company was recently quoted an annual premium of \$400,000 for total coverage of \$1 million.

Part of the difficulty is the recent inflation of product liability claims, and of insurance premiums. The insurance industry is also under pressure from the number of product liability suits in federal courts, up by 750 per cent since 1974, and the average jury award, which has risen from \$400,000 to more than \$1.8 million during the same period, according to the American Insurance Association.

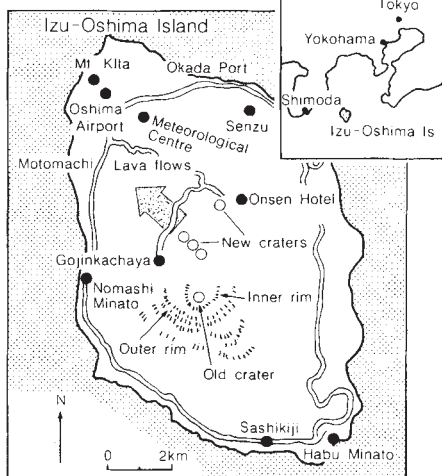
Insurers are especially fearful of the doctrine, recently introduced in California, of "joint and several liability", which means that a party only partly at fault can be obliged to become the "deep pocket" from which a whole claim must be met. Chemical, pharmaceutical, aircraft and car manufacturers are most at risk.

This, say insurance companies, is why they prefer to stick to the devils they know, such as automobile insurance and workers' compensation, than to take on unknown quantities such as biotechnology. They fear that huge claims may be made in the light of future knowledge, as they have been for medical products, when initial ignorance of harmful effects may not be a defence.

The ABC scheme was recommended by a study conducted by insurance brokers Johnson and Higgins of New York. Twenty-four companies from around the

world, including well-known companies such as Celltech and Charles River, have chipped in enough to establish the biotechnology insurance cooperative legally in the state of Vermont; it should be ready for "activation" by early December. Membership will be open to any company in the field, but 30 members will be required to make it viable. The Industrial Biotechnology Association, which represents generally larger companies, has considered and rejected such a scheme, as many of its members are able to carry their own insurance.

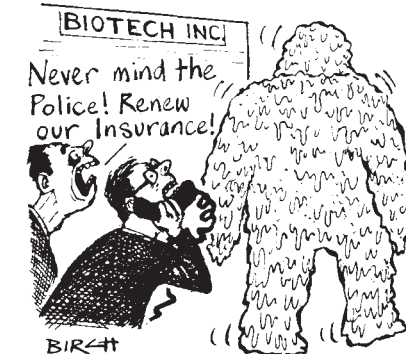
The ABC captive insurer will provide only modest coverage at the outset: \$1 million of either primary or "excess" coverage, with an aggregate ceiling of \$3 million. But coverage would increase if the scheme expands; 44 companies have asked for quotes. ABC hopes the scheme will include insurance against environmental damage for such products as mic-



that typically precedes major eruptions as magma rises to the surface. And at the end of October the Volcanic Eruptions Prediction Liaison Council declared that there were no signs that a large-scale eruption was imminent.

Even after eruptions began on 15 November, the council confidently predicted that lava flows would be confined to the caldera by the outer rim of the volcano, and Oshima enjoyed a temporary boom as hundreds of tourists flocked to the island each day to enjoy the spectacular eruption and lava flows.

After 10 million cubic metres of lava had spilled over the inner rim of the volcano and the rivulets of lava in the caldera had frozen in their path, the Meteorological Agency announced that activity appeared to be decreasing and the island's



robial pesticides, but admits that may be expensive. All participants will be inspected by risk management experts to ensure they are "doing things by the book".

The American Insurance Association is generally sceptical of captive insurers, saying that insurance costs reflect the costs of meeting claims, no matter who does the underwriting. The insurance industry says that the provision of medical malpractice insurance through similar "risk retention groups" is now as expensive as through commercial agents.

Tim Beardsley

CNRS

Union power may not erode

THE 20,000 staff, £600-million budget French national research council, CNRS, is to be given a lighter management structure and put back in the hands of the 'best' scientists after years of technocratic rule from the top and foot-dragging by trades unions.

That at least is the theory now being put out by the new CNRS administration, encouraged by the French prime minister, Jacques Chirac. Last week Chirac and his ministers gave approval for a new de-unionized structure for the research council's elected evaluation committee, the 1,000-strong Comité National. This gives back to CNRS power lost earlier in the year when the previous Comité National was declared illegal on a technicality, and clarifies at least part of the new government's policy towards science.

According to Serge Feneuille, director-general of CNRS, one of the new slogans will be that "science cannot be unionized", and the restructuring of the Comité National confirms it. In the elections, which will probably be complete by April 1987, at least half the candidates, especially the most senior, will no longer have to declare a union affiliation but may stand under their own name. Places on the committee for the engineers, technicians and administrators who are most unionized are to fall.

But although these developments are meant to loosen the hold of the unions over the system of promotions and grants, the new structure, according to its opponents, will shift power from the unions to another sector that could be equally conservative, university lecturers who do no research, but who will now be free to vote and stand for election. This move is a concession to university opinion that would have preferred CNRS to be dismantled. The resulting compromise, say critics, will merely hamstring the scientists in a new way. In the life sciences, many non-researching physicians from teaching hospitals will be joining the evaluation committees to make their opinions of French biology, and biologists, felt.

Whatever the politics of the Comité, there is some relief among 'les admissibles', the 512 young hopefuls who were appointed to full-time life-long CNRS research positions earlier this year by the previous Comité National, only to have the Comité declared illegal. At least now there is a legal Comité to look into their case, but, according to CNRS, they will have to take their examinations again. The same number of posts, 512, will be made available, and the resulting appointments will be counted as 1986 appointments with no bearing on the level of further recruitment in 1987. The proper

1987 appointments round (some 400–450 posts) will then take place in September and October, with a return to a normal spring timetable in 1988.

Meanwhile, Feneuille is known to have considerable interest in the introduction of short-term postdoctoral positions into the French research system which now

Soviet academy

Marchuk in ministerial mode

THE Soviet Academy of Sciences, already regarded by Soviet planners as the chief coordinator of all pure and applied research in the Soviet Union, is to acquire a new role. Addressing the Supreme Soviet (the Soviet Union's bicameral parliament) last week, Guri Marchuk, the new president of the academy, said that it is to start producing forecasts for the major areas of science so as to identify the parts of the national economy ripe for development.

The new role of forecasting is part of a major reorganization of the structure and management system in line with the restructuring of the whole Soviet economy launched by Mr Mikhail Gorbachev last year. Under the new scheme, Marchuk said, the "specialized departments" of the academy will play a larger part, so as to direct scientific institutions towards the most important problems of fundamental and applied science.

How these changes will work is not yet clear. The major innovation so far of the Gorbachev era in science and technology, the "interbranch scientific-technical complexes" (MNTKs) are, in Marchuk's words "producing stress and difficulties". These organizations were supposed to "accelerate scientific and technical progress" by cutting across the interdepartmental bureaucracy and the barriers between science and industry. (At present the gap between initial breakthrough and large-scale production is said to be as much as 11 years.)

But as yet, the only MNTK to have received official commendation is the Paton Institute of Electric Welding of the Ukrainian Academy of Sciences, which has existed for many years and which served as the model for the MNTK concept.

Marchuk also repeated in his speech the targets outlined in the draft plan for 1987 (computers — both main-frame and personal; plasma, radiation and laser treatment of specialized materials; powder metallurgy; and biotechnology) and the Comecon Comprehensive Programme for Scientific and Technical Progress up to the year 2000. But, he warned, if the Soviet Union's "fine scientific and technical foundation for attaining the highest world

locks itself into the appointment of a scientist for life at the age of 26. Feneuille believes that French industry, in which science is severely under-represented is in need of 'irrigation' with scientists. He argues that a system of postdoctoral appointments to CNRS in excess of the long-term need would allow the scientific irrigation of industry, as happens regularly in the United States and Britain. Such a change, however, could be political dynamite in France.

Robert Walgate

standards" is to be properly employed, the planners, and in particular those involved in the capital construction of new enterprises, must play their part.

This is a new line. Hitherto, scientists have blamed industrialists for not using their results; more recently industrialists have begun to retaliate, saying that the results they were offered were unsuitable in practice. Marchuk now seems to be arguing that scientists can produce what industry needs only if the planners properly specify what is required.

But the real novelty of Marchuk's speech was not so much what he said but the fact that he addressed the Supreme Soviet in a virtually ministerial manner. As former head of the State Commission for Science, of course, he is no stranger to the Supreme Soviet. But the fact that he would have addressed it in this way at this early stage suggests that his presidency of the academy will be more vigorous than that of his predecessor.

Vera Rich

Networking the begging-bowl

Strasbourg

THE general assembly of the European Science Foundation (ESF), the most penniless and self-effacing of European science institutions, came to no particularly decisive conclusions last week. But the foundation did agree that it would ask its members (emphatically not governments but the research councils that governments support) for extra funds to keep its network programme alive.

That programme springs from an inquiry by ESF two years ago, at the behest of the European Commission, which showed that European researchers were too often isolated from each other. Largely because of a donation of a million French francs, the network programme last year brought 400 people together at 35 international meetings. The flavour, this year, is polar research, on which everybody is keen. But there are no funds, whence the begging bowl.

Robert Walgate

Technology database

Best buys in UK research

EVERY researcher in Britain is to be mailed on behalf of an on-line computer database, designed to sell British talents to industry worldwide. Called British Expertise in Science and Technology (BEST), the commercially run database has been endorsed by the British government and the Science and Engineering Research Council (SERC).

Researchers in universities, polytechnics and the research councils are being asked to take part in the programme, in part intended to provide the supposed missing link between researchers and industrialists. Thirteen thousand records on researchers and their work now constitute the core of the database. A typical record will show a researcher's qualifications, publications, past and present research work, specialities, peripheral interests, industrial experience, current and previous positions. With the appropriate telecommunications link, the database can be tapped from any part of the world.

The database is part of the response by British universities to the publication in June 1983 of the report by the government's Advisory Council for Applied Research and Development (ACARD), which urged closer links on research between higher education and industry. In 1984, the universities set up a committee representing government departments and agencies as well universities and polytechnics. The committee selected the best proposal by Longman Cartermill, part of the Longman publishing house, from among other bids submitted.

Since the launch of the database in the spring, the operators claim to have attracted more than 100 industrial subscribers. Data entries to the system are free. Revenue will derive from subscribers, who are charged £200 for a single search or fees ranging up to £10,000 a year for full corporate use.

The terms of an agreement between SERC and the operators have yet to be concluded, but to start with, SERC will provide information on its research grants in return for access to the database for its own purposes.

Bill Johnstone

• Meanwhile, the British Library is to launch its separate on-line computer database this week. That will focus more on research projects than researchers. The database contains 65,000 records of research being conducted in Britain. The system, called Current Research in Britain (CRIB), has evolved from an annual printed publication produced by the library. It has a £50-an-hour connection charge and is operated for the British Library by Pergamon InfoLine. □

Information technology

Filling the post-Alvey gap

A £1,000 million five-year plan to strengthen Britain's research in and application of information technology was proposed last week by the IT86 committee, set up this spring under Sir Austin Bide, the now-retired chairman of the Glaxo pharmaceutical group. The request for funds on this unprecedented scale is meant to continue the five-year Alvey programme of research and development in information technology, whose £350 million is now almost fully committed.

More than 300 projects have been approved under the Alvey programme, but the Bide committee considers it too soon to evaluate their success except to say that they have been successful in drawing industry and universities together more closely "than was previously the case". The committee also says the programme has stimulated collaboration between companies normally in competition.

Of the total £1,000 million, the committee proposes that roughly £425 million should come from the British government and the remainder from industry. The government contributed £200 million to the total cost of Alvey projects.

The Bide committee estimates that the research element in the continuation programme will cost £550 million, of which the government is expected to contribute £300 million. As under Alvey, academic research would be fully supported by government, but research by industrial companies receives only half-support.

The committee also endorses some of the conclusions of a report by the British government's Advisory Council for Applied Research and Development (ACARD) which, earlier this year, was critical of the British software industry.

That document forecast a continued increase in the British trade deficit in information technology.

Much of the Bide committee's emphasis is on the need for converting research into marketable products. The committee says that applications schemes must involve collaboration between researchers and industry and that academic contributions to joint ventures should "go well beyond" traditional information-technology departments. The report also includes a long shopping-list of projects to be supported. Among applications projects, it singles out security systems, clinical data processing, interactive distance learning aids, distance learning for management, manufacturing control, electronic funds transfer at point of sale and software for railway signalling.

On research, the Bide committee recommends a number of areas to which effort should be directed, including systems design and management and new programming tools. Not all the support now requested need be new money; part of it may come from the funds already earmarked for the second phase of the European Strategic Programme for Research and Development in Information Technology (ESPRIT).

No doubt in recognition of the difficulty of extracting large amounts of money from the budgets of individual government departments, the committee says that the funds should come not only from the original sponsors of the Alvey programme (the Departments of Trade and Industry and of Defence and the Science and Engineering Research Council) but from other departments with an interest in the applications suggested, either as sponsors or as users.

Bill Johnstone

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

The Straits of Gibraltar, with Spain to the left and North Africa to the right. On page 313 of this issue we publish a 20-page supplement on science in the Iberian peninsula. Photo: Oblique view from Skylab, courtesy NASA.

Dissent disallowed in Singapore

SIR—There exists an interesting juxtaposition of events in Singapore this week. On the one hand, scientists from all over the world are gathering for the Fourth Federation of Asian and Oceania Biochemists (FAOB) meeting, to present and discuss work freely. This meeting is important, as it is timely, to the powers that be in Singapore, in that it precedes the opening of the new Institute for Molecular and Cell Biology there, for which the talents of the international scientific community are being anxiously courted. In stark contrast, the voice of dissent in Singapore's parliament, Mr Joshua Jeyaretnam, is effectively silenced this week; together with the chairman of the opposition party, Mr Wong Hong Toy, he finds himself in the cells of Singapore's Changi Prison.

The history of events that have led to fines, imprisonment and the removal of the leader of the opposition from his seat in parliament is adequately dealt with in such publications as *Time Magazine* (8 September 1986), *Financial Times* (12 November 1986), *Far Eastern Economic Review* (4 September and 20 November 1986) and the *Asian Wall Street Journal*. As a result of these articles, in which the motives of the government and the independence of the judiciary are called into question, two of these publications have incurred sanctions. The net overall effect of these events is to suggest that the voice of dissent in Singapore, be it from individuals, editors or members of parliament, is to be silenced at any cost. It is against this political backdrop that the Singapore government, advised by a number of eminent scientists from the United Kingdom, the United States of America and West Germany, seeks aggressively to recruit internationally scientists upon whose cooperation the success of the FAOB meeting and that of the Institute for Molecular and Cell Biology rests.

This situation, taken as a whole, raises some quite difficult questions, addressed especially to those advising or considering working at the Institute for Molecular and Cell Biology in Singapore. When a scientist enters into a commitment of research in Singapore, does he retain his right to comment on what he sees? If the independence of the judiciary and the freedom of the press are called into question, who is left to take the responsibility of safeguarding the rights of the individual to criticize? This responsibility sometimes comes to rest in unexpected quarters. In this case, because their services are of value to the Singapore government, some of this responsibility lies squarely with the international scientific community. It is up to us to satisfy ourselves (the facts are on record) that the individual in Singapore who wishes to voice legitimate criticism

has the freedom to do so. If this freedom does not prevail, then the scientific community at large, being in a position of influence, must not flinch from the responsibilities associated with this position, and must use this influence to foster the necessary change of attitude within the government in Singapore.

No government should assume that it can, through institutions such as the Institute for Molecular and Cell Biology in Singapore, exploit the talents of the international scientific community and expect the scientists concerned to turn a blind eye to the fate of the people in whose country they work.

LOUIS MAHADEVAN

London NW6, UK

Fluoridation

SIR—The 1983 Anglesey study¹ does not demonstrate the effectiveness of fluoridation, as Francis B. Reed claims². Apart from its questionable late selection of a known high-carries non-fluoridated control population different from that originally planned, without pre-fluoridation information on the populations being compared — making its “strictly blind conditions”² worthless — its concluding assertion of “the universal finding that caries experience in a fluoridated community is consistently lower than in neighbouring non-fluoridated communities” is shown to be untrue by the many studies cited in Diesendorf's Commentary³. In New Zealand we have dental health figures for our entire child population, not just for samples as in “fluoridation trials”. The figures for Auckland show that the children in the non-fluoridated part of the city require fewer fillings, not more⁴. The non-fluoridated off-shore island children also need fewer fillings, just like the fluoridated island of Anglesey.

JOHN COLQUHOUN

Education Department,
University of Auckland,
Private Bag, Auckland,
New Zealand

1. Jackson, D. *et al. Br. Dent. J.* **138**, 165–171 (1975); **158**, 45–49 (1985).
2. Reed, F.B. *Nature* **323**, 198 (1986).
3. Diesendorf, M. *Nature* **322**, 125–129 (1986).
4. Colquhoun, J. *Community Dent. Oral Epidemiol.* **13**, 37–41 (1985).

UK performance

SIR—Your contributors have recently reviewed two influential reports on British performance in research^{1,2}. One concludes that “the main findings are that in basic research in science as a whole, UK performance has deteriorated since the early 1970s both in absolute terms and relative to other countries, notably FRG and

Japan”. The figures that most strikingly emphasize this point are the citation rates.

This broadbrush conclusion disguises a more precise relationship. The relative proportions of each of the competitors' total research budgets spent on defence in the early 1980s have respectively been 0.5, 0.2 and 0.02, demonstrating a clear negative correlation between the proportion of research funding for defence and the corresponding civil research performance.

One can only guess why this should be so, if indeed the relationship is real at all, let alone representative of the world at large, derived as it is from only three (very important) examples. Yet if it is real it smacks of a sort of logic: the greater the proportion spent on defence, the greater the number of defence scientists required to spend it and consequently, from a limited resource, the smaller the number in civil research; also the greater the proportion spent on defence, the greater is the restriction on research reports available for citation. The former points towards a real problem for UK civil research; the latter to a possible mitigating explanation of its apparent recent lacklustre performance.

BRUCE DENNESS

Bureau of Applied Sciences,
Wydcombe Manor,
Whitwell,
Isle of Wight PO38 2NY, UK

1. *Evaluation of National Performance in Basic Research*, ABRC Science Policy Studies No. 1 (Advisory Board for the Research Councils, London, 1986).
2. Martin, B.R., Irvine, J. & Minchin, N. *An International Comparison of Government Funding of Academic and Academically Related Research* – Vols 1 & 2, ABRC Science Policy Studies No. 2 (Advisory Board for the Research Councils, London, 1986).

Backfiring boycott

SIR—The signatories of the Strategic Defense Initiative (SDI) boycott pledge (see *Nature* **323**, 747; 1986) may find themselves in an embarrassing position, should the United States and the Soviet Union agree on how to regulate laboratory research on SDI. Would they recant their position, and on what grounds? The research programme (and that is all that the academics are involved in) will certainly not become technically any less dubious simply by receiving Mr Gorbachev's blessing. And were the Western scientists to continue their boycott unilaterally, they would simply be paving the way to a technological breakthrough by the Soviets, tempting them to break out from the Anti-Ballistic Missile Treaty. Is this course of action any less misguided than participation in SDI research? The most disturbing feature of the anti-SDI pledges circulating in the United States and now in Europe is that they are black-and-white statements that do not allow for any shades of grey.

NICHOLAS ZUMBULYADIS

686 Seneca Parkway,
Rochester, New York 14613, USA

How to make antimatter last

An experiment at CERN has shown that antiprotons can be kept for several minutes in an electromagnetic trap. But the prospect of making reality of the science fiction remains remote.

SCIENCE fiction writers, ingenious people as they are, include among the standard props of their trade a ready supply of antimatter, the most portable of all high-explosives. During the past few years, there has also been a disturbing trickle of accounts in the popular press suggesting that the world's accelerator laboratories are hard at work on the development of techniques for making antimatter for use in military operations. A glance at one of the first published accounts of success in storing antiprotons (*Phys. Rev. Lett.* **57**, 2504; 1986) should persuade all those concerned that it will be some time before the fiction can become reality.

For other reasons, the new development is interesting and important. Working at CERN, the European high-energy physics laboratory at Geneva, a group from the University of Washington at Seattle (Gabrielse, Kei, Helmerson, Rolston, Tjoekler and Trainor), two people from the University of Mainz (Kalinowsky and Haas) and one from the Fermi National Accelerator Laboratory (Wells) have been able to decelerate antiprotons to the point at which their energy is sufficiently small to allow them to be captured in an electromagnetic trap. The experiment is a technical achievement in its own right, but the prospect that it will now be possible to study antiprotons essentially at rest, and stable for some minutes, raises the prospect of a series of previously impossible experiments.

CERN's advantage in this connection is the facility called the Low Energy Antiproton Ring (LEAR), which is a device for storing bunches of antiprotons produced by the Super Proton Synchrotron (as modified so that protons and antiprotons travel in opposite directions) and reducing their individual energy by four orders of magnitude to 5 MeV or thereabouts. Novel experiments are possible even at this energy (see Batty, C.J., *News & Views*, *Nature* **323**, 487; 1986). But even 5 million volts is far greater than the electrostatic potential well of the kinds of traps that might be constructed in the laboratory.

The obvious way to make possible the trapping of antiprotons would be a further elaboration of the LEAR storage ring, but although CERN has ambitions in that direction, the extra equipment has not yet been built. Accordingly, the Washington-Mainz group has fallen back on the simplest of all devices for robbing fast charged

particles of energy — that of making them traverse a dense target of matter, a solid beryllium window. In practice, the experiment uses virtually instantaneous bunches of antiprotons, each of them containing between 10 million and 1,000 million particles.

Inevitably, the process is inefficient. With the thickness of the window chosen to be the range of antiprotons, at the energy discharged from LEAR, in beryllium, only about 1 in 10,000 of the original bunch emerges from the target with an energy less than 3 keV, which means that typical bunches leave only some 10,000 antiprotons for catching in a trap.

That consists of two components, of which the simplest is a powerful axial magnetic field to prevent slow antiprotons moving too quickly away from the axis of the experiment. The other is a conceptually simple but technically sophisticated pair of electrodes, each with a hole in the centre (a third of a centimetre across) to allow the passage of the antiproton beam, and roughly a dozen centimetres apart. Before the arrival of an antiproton bunch, the first electrode is at earth potential and the second, or more distant, has a potential of -3 kV so as to repel antiprotons less energetic than 3 keV. But while the bunch is travelling the length of the trap, the voltage of the first electrode is switched, in a matter of nanoseconds, to match the potential of the first.

For the time being, the objective is to see how efficient such a simple trapping technique may be. As these things go, the mechanism seems to work as well as anybody expected in advance. The population of antiprotons can be measured at any time by abolishing the repulsive potential on the more distant of the two electrodes, whereupon the antiprotons are guided by the axial magnetic field to a target where they annihilate, yielding pions which can be counted. Apparently, the counting equipment is not yet fast enough to measure the population of a newly filled trap, but typical records show that, after a millisecond, the trap may contain 300 or so antiprotons. One of the published records shows that after storage for 100 seconds, it was still possible to record 31 antiprotons, spread over a mere 5 microseconds. On balance, it seems as if it is possible to catch roughly 3 in a million of all the particles discharged as a bunch from LEAR.

At this rate, sheer numbers are less important than the other properties of the

trap, which will determine what use can be made of a supply of essentially static antiprotons. The authors of the experiment brim over with suggestions for what may happen next. One scheme is to make single atoms of anti-hydrogen, perhaps by building two traps alongside each other, one for antiprotons and one for positrons. Firing a beam of positronium atoms (pairs of positive and negative electrons) at antiprotons in a trap might be a simpler way of accomplishing the same goal. The prospect of precision measurements of quantities such as the antiproton mass, or even of the gravitational acceleration of antiprotons, will intrigue many observers of this new development.

Technically, the improvement of the present means of trapping antiprotons is an obvious line of evolution, but the authors write of the still more promising line of development in which antiprotons trapped as they describe would be transferred, bunch by bunch, to a trap whose geometry is not constrained by the need to capture as many particles as possible from LEAR. Such a device would therefore be better suited for keeping even these unstable particles for long periods of time, a day or even longer. That is what will have to be done if some of the more interesting measurements now suggested are to be carried out.

All this is of course a long way from what the science fiction writers require of antimatter. While the best that people can do is to store a few hundred antiprotons for an hour or so, the prospect of antimatter weapons will remain very distant. The essential difficulty is a score or so of orders of magnitude. True, the annihilation of an antiproton by, say, a proton of the kind that turns up in ordinary matter is more energetic, by a factor of say a hundred, than the fission of a uranium nucleus. But Avogadro's number (6.10^{23}) works emphatically in the other direction. So long as numbers of antiprotons that people can collect in antiproton traps are small enough that it is feasible to think of counting them, the popular newspapers' accounts of what high-energy physics is really for will remain in the realm of fantasy. It is nevertheless outrageous that there have been in the past few years, such a torrent of tales along these lines. They will always be fairy stories. So, too, perhaps sadly, will be the relevant parts of science fiction.

John Maddox

Evolution

Contemplating life without sex

John Maynard Smith

FROM the selfish-genetic point of view, there are obvious advantages to parthenogenesis. A gene that suppressed meiosis in female diploids would be transmitted to all female offspring instead of only to half and thus would, all else being equal, double in frequency in each generation. Why then, does parthenogenesis not replace sex? The answer has to be that all else is not equal. Why not? One good reason for studying parthenogens is for the light they may shed on the evolution of sex, and the problems of parthenogens were the topic of a recent symposium* from which it is clear that for one reason or another it is very difficult to give up sex once you have it.

Most existing parthenogens have evolved from sexual forms and are hampered to varying degrees by the persistence of mechanisms appropriate only to their ancestral lifestyle. As so often, the fruitfly *Drosophila* provides a good example. Females of many species lay an occasional egg that develops without fertilization. But it is a long way from the occasional egg to functional parthenogenesis (A. Templeton, Washington University, St Louis). Parthenogenetic laboratory strains now exist in several species, including *D. melanogaster*, but it took much selection to establish them. Before selection, unmated females lay few eggs, and even fewer start development. Of those that do start, most die because of genetic homozygosity resulting directly from the mechanism of parthenogenesis. The mechanism is known as automixis: the egg undergoes meiosis, and diploidy is restored by the fusion of two of the four products. Because natural populations carry many deleterious recessive genes, automixis results in heavy mortality. The only species which is parthenogenetic in nature, *D. mangabeiri*, has overcome this problem in an ingenious way. First, the two haploid nuclei that fuse are those that are separated at the first meiotic division; second, there is no crossing over. Together these devices ensure that the offspring are genetically identical to the parent: heterozygosity is preserved. Crossing over is prevented by inversions.

The picture that emerges from these studies of *Drosophila* is probably typical. The disadvantages suffered by recently arisen parthenogens are the direct result of their sexual ancestry; but these 'sexual hang-ups' can be overcome by selection. The precise means of overcoming them varies. Some parthenogenetic animals, and many plants, opt for abandoning

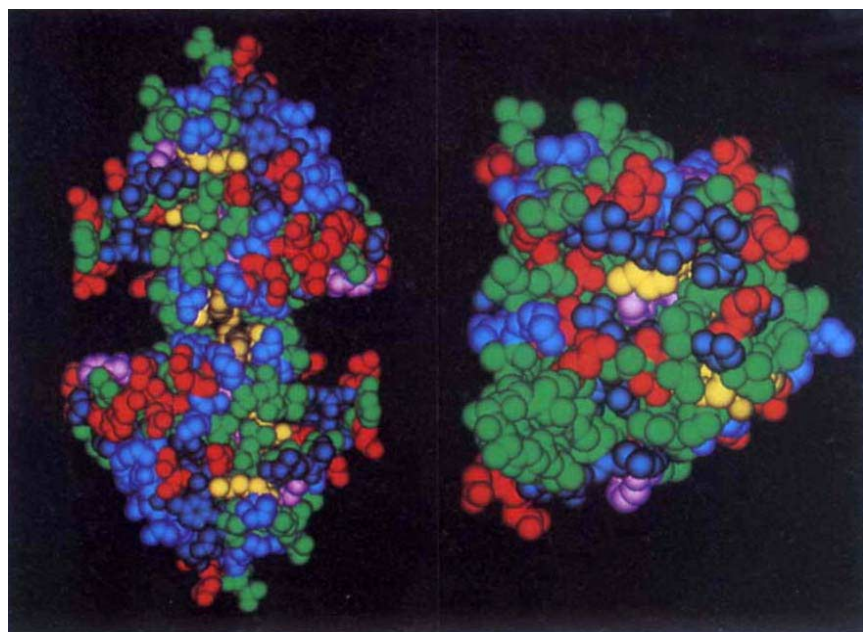
meiosis altogether — apomixis; this is the pattern in the cyclically parthenogenetic aphids and cladocerans. Alternatively, in the parthenogenetic lizards of the genus *Cnemidophorus*, there is a premeiotic doubling of the chromosomes, followed by the pairing of sister chromosomes in the first meiotic division. This has the same genetic consequence as the meiosis of *D. mangabeiri*: offspring are genetically identical to their parents, and do not suffer the enforced genetic homozygosity that is the usual result of automictic parthenogenesis. The parthenogenetic species *C. uniparens*, which is a highly heterozygous species hybrid, consists of a single clone: members of the species accept scale grafts from each other.

Another common type of sexual hang-up in parthenogenesis is the persistence of the need for sperm to trigger development — pseudogamy (that is, the stimulation of egg development by a sperm whose chromosomes do not participate in forming the new individual).

In some lower vertebrates (discussed by H. Hotz and T. Uzell, University of

Illinois and R.C. Vrijenhoek, Rutgers University), the sperm is probably needed to provide a centriole. (No such need, however, exists in lizards.) Pseudogamy is not uncommon in insects. In plants, a nucleus from the pollen tube may be needed to fertilize the endosperm, even if the embryo is parthenogenetic. Some parthenogenetic frogs (*Rana*) and fishes (*Poeciliopsis*) have evolved what is perhaps the strangest mechanism of all for dealing with this requirement. It is known as hybridogenesis. *Rana esculenta*, for example, is a hybrid between two sexual species, *R. ridibunda* and *R. lessonae*. In the germ line, *lessonae* chromosomes are thrown out, so that the egg contains only the *ridibunda* component. If, as is typical, the hybrid female mates with a *lessonae* male, the hybrid *R. esculenta* is reconstituted. Clearly, pseudogamous and hybridogenetic forms cannot wholly replace their sexual ancestors. In some major taxa, for example mammals and gymnosperms, parthenogenesis has not been reliably recorded, but the nature of the hang-up is not known.

Despite the widespread occurrence of sexual hang-ups there are many parthenogens that seem to have escaped the disadvantages derived from their sexual ancestors. The taxonomic distribution of parthenogens, however, suggests that they may be successful in the short run,



THESE three-dimensional views of β -lactoglobulin show two identical monomers tenuously linked. Each monomer is similar to plasma retinol-binding protein and each contains a non-polar pocket similar to that occupied by retinol in retinol-binding protein. The non-polar pocket is in the centre of view in the right-hand figure. The function of β -lactoglobulin, a milk protein, has long been unknown; these new models suggest that it is a transport protein. The structure is described in more detail by M.Z. Papiz and colleagues from the Universities of Leeds, Edinburgh and Uppsala, on pages 383–385 of this issue. The colours represent different types of amino-acid residue: red, acidic; dark blue, basic; light blue, polar; purple, aromatic; yellow, sulphur-containing; green, non-polar. Computer program by Dr H. Nakamura. □

*A meeting on parthenogenesis organized by L. Kirkendall and N.C. Stenseth at Finse, Norway, 17–22 September 1986.

but are short-lived on an evolutionary timescale. Thus, most parthenogens have close sexual relatives. It is rare for a taxon higher than a species to consist wholly of obligate parthenogens. The parthenogenetic order Bdelloidea of rotifers, and the tribe Traminini of aphids, are important exceptions that challenge further research.

Among higher plants, taxa such as *Taraxacum* (dandelions) and *Rubus* (blackberries) that contain many parthenogens also contain sexual species, and new clones are continuously arising. Two theoretical reasons have been suggested for this short evolutionary life expectation. First, evolution to meet changed circumstances will be slow in the absence of genetic recombination. Second, there may be a tendency to accumulate slightly deleterious mutations (Muller's ratchet¹). Although the theory is clear, it is hard to provide empirical support. A major difficulty lies in deciding how old particular clones may be, and in distinguishing between variation that has arisen by evolution within a clone and that which reflects multiple origins from sexual ancestors. No-one was prepared to make even a rough estimate of the age of any clones.

The simplest picture is as follows. Most parthenogenetic 'mutants' fail to establish themselves because of various sexual hang-ups. Some, under intense selection, overcome these difficulties and achieve short-term ecological success. Despite their twofold reproductive advantage, however, success is short-lived on an evolutionary timescale because of their relative inability to adapt, and perhaps because of the accumulation of mutations.

Is any further explanation needed? Williams² argued that it is, essentially because the coexistence of closely related sexual and asexual populations implies that sex must confer some immediate advantage. He thinks that this advantage arose from the genetic variability associated with sexual reproduction. This can take two forms: first, a sexual population will be more variable than an asexual one; and second, the offspring of a single sexual female will vary more than the offspring of an asexual one.

The variability of a sexual population may be relevant in the following way. A single clone, with a single genotype, is unlikely to displace its sexual ancestor over its whole ecological and geographical range. Vrijenhoek has called this the frozen niche hypothesis (Bell's³ tangled bank model is, I think, similar), and argues that, as the number of different clones increases, the relative success of the sexual species should decline. He finds that, in *Poeciliopsis*, clones do differ physiologically and in their ecological adaptations, and that the proportion of the population that is sexual declines as the number of distinct clones in a stream decreases. In *Poeciliopsis*, however, parthenogens need

to be mated, so they cannot wholly replace their sexual ancestors. Clonal diversity is now being studied, using biochemical markers (and, in vertebrates, immunological methods), in many plants and animals, and this should help us to decide on the importance of population variability in conferring an immediate advantage on sexual populations.

The Williams 'balance' argument may also be relevant to cyclical parthenogens such as cladocerans (M. Lynch, University of Illinois) and aphids (N. Moran, University of Arizona). We must be cautious, however, because there is often an ecological difference between the sexually and asexually produced eggs. For example, in *Daphnia* the resistant winter eggs are often produced sexually. If this was a universal rule, the maintenance of sex could be explained by the need to produce winter eggs. But there are obligately parthenogenetic strains of *Daphnia*, some of which produce winter eggs asexually. The situation is similar but more complex in aphids. Most aphids have two plant hosts, and the sexual forms attack the secondary host. No obligately parthenogenetic aphid has retained the ability to utilize two hosts, but there are many cyclically parthenogenetic aphids that are confined to a single host, and these have given rise to many obligately parthenogenetic strains. Some modification of the frozen niche hypothesis, therefore, may be needed to explain why all cladocerans, and aphids with only one host, do not wholly abandon sex.

Williams argues that genetic variation among the offspring of a single female may also be important. He uses the famous analogy that a parthenogenetic female is like a man who buys 100 tickets in a lottery and finds that they all have the same number. I have argued⁴ that this model only works if the offspring of one female compete with one another, but showed that, provided that there is sib competition, the model can give rise to a twofold advantage to sex.

But do we need to take into account the genetic variability of individual families, or are the other explanations outlined above sufficient to explain the failure of parthenogens to replace sexual forms? There is one context in which within-family variability may be relevant.

Consider a population of diploid parthenogenetic females; that is, females that produce some of their offspring parthenogenetically and some sexually. Why does such a population not evolve towards obligate parthenogenesis? From the point of view of its own proliferation, there would be an intrinsic advantage to a gene that suppressed meiosis, and it is hard to believe that there would not be some genetic variance affecting the proportion of offspring produced asexually and on which natural selection could work. We

cannot appeal to the greater variability of the sexual population, because the range of genetic variability is the same in the sexually and asexually produced offspring. The sib competition model might, therefore, be necessary.

Do such facultative parthenogens exist? In animals, I know of no clear example. Populations with a female-biased sex ratio usually consist of a mixture of obligate parthenogens and a bisexual population. But there are many cases that merit further investigation. In plants, there do seem to be examples. One, the diploid form of *Potentilla argentea*, was described (S. Asker, Lund). These cases need genetic and ecological study.

I formed the strong impression that the study of parthenogenesis is, at last, beginning to tell us something about the evolutionary significance of sex. Much remains to be done. The Bdelloids remain something of an evolutionary scandal. We need to know the age of at least some natural clones; the extent of genetic diversity that can arise within them; and whether deleterious mutations accumulate in them by the ratchet mechanism. Further studies of clonal diversity will show how important the frozen niche and tangled bank models in fact are. The suggestion^{5,6} that sex is needed in the evolutionary race against parasites should be investigated. The few clear cases of diploid facultative parthenogenesis should tell us whether sib competition is relevant. I think that the problem of the maintenance of sex in higher eukaryotes is different both from the problem of the evolution of recombination in prokaryotes and from the origin of meiosis and syngamy in single-celled eukaryotes. The three problems may have different solutions. Indeed, I am reasonably sure that the first two do have different solutions, because the primary function of recombination in prokaryotes is probably the repair of damaged DNA. Although DNA repair is also needed in eukaryotes, I do not think that the suggestion⁷ that meiosis and syngamy serve a repair function is tenable: indeed, one of the reasons why parthenogens are interesting is that they demonstrate that this suggestion is mistaken. As far as the maintenance of sex in organisms like ourselves is concerned, I think the main theoretical explanations have already been formulated, and our task is to decide on their relative importance. □

1. Muller, H.J. *Mutat. Res.* 1, 2 (1964).
2. Williams, G.C. *Sex and Evolution* (Princeton University Press, 1975).
3. Bell, G. *The Masterpiece of Nature* (University of California Press, Berkeley, 1982).
4. Maynard Smith, J. *J. theor. Biol.* 63, 245 (1976).
5. Jaenike, J. *Evol. Theory* 3, 191 (1978).
6. Hamilton, W.D. *Oikos* 35, 282 (1980).
7. Bernstein, H., Byerly, H.C., Hopf, F.A. & Michod, R.E. *Science* 229, 1277 (1985).

John Maynard Smith is Professor of Biology at the University of Sussex, Falmer, Brighton BN1 9QG, UK.

Materials science

Nuts to tough ceramics

Paul Calvert

MACADAMIA or Queensland nuts are much used in high-class cookies but are very difficult to remove from their shells without special equipment. In a recent article, J.S. Jennings and N.H. Macmillan (*J. Mater. Sci.* **21**, 1517; 1986) discuss the properties of the fibrous macadamia nut-shell and shed some light on the prospects for fibre-reinforced ceramic materials which are currently thought to be the route to tough ceramics for use in engine components.

Jennings and Macmillan looked at the microstructure of the nutshells and measured the mechanical properties of various sections of the shells in the wet and dry states. It is very difficult to obtain reproducible mechanical measurements on biological materials because the properties are frequently dependent on the age and moisture content of the tissue and vary with orientation and position in the organism as well as being vulnerable to flaws. This creates something of a philosophical problem for engineers who are faced with a natural system, when they are more used to uniform properties measured on large pieces of material. In this (nut) case the properties are surprisingly uniform and are best described as those of 'isotropic wood'. That is, the moduli, strength and toughness are comparable with the geometric averages of those parallel to the grain and across the grain of a softwood. A nice recent example of this grain effect is a study of kelp (Vincent, J.F.V. and Gravell, K. *J. Mater. Sci. Lett.* **5**, 353; 1986), which forms 'leaves' by the tearing action of the sea along the weak direction parallel to the fibres. The work of fracture decreases by 500 times from fracture across the fibres to fracture parallel to the fibres.

The Young's modulus (elasticity in tension) of the nutshells is about 5 GPa, where wood is about 10 GPa parallel to the grain and 0.5 GPa across the grain. The (fracture) strength of the shells is 50–80 MPa, where wood is about 100 MPa along the grain and 5 MPa across. In terms of a complete nut this corresponds to a hungry gourmet needing to exert about 50 kilos of force on a normal pair of nut-crackers. The fracture toughness is calculated from the stress needed to propagate a crack in a notched sample, and is thus a measure of one's ability to open the nut by making a small cut and levering on it. This value is in the range from 0.8–2 MPa m^{1/2} for the nutshell whereas wood is 1 and 2 MPa m^{1/2} parallel and perpendicular to the grain, respectively. As Jennings and Macmillan point out, these properties are

not terrific when one takes into account the fact that the density of nutshells is 2.5–3 times that of wood. The advantage of this design seems to lie in the isotropic properties which ensure that the shell is not weak along any particular direction and so cannot be split like wood, on the same principle that plywood gains strength by having layers with crossed grains. This toughness precludes attacking the nut by a shrewd wallop with a hammer and chisel.

The shell is made up of fibrous cells, each about 30 µm in diameter, which are arranged in bundles of tens or hundreds of parallel rods. The bundles then pack to fill the volume. When the shell breaks, the fracture either goes through the rods or around them, depending on whether the crack is perpendicular or parallel to the rods. The toughness then arises from the need for the crack to change direction as it traverses the shell, possibly with some microcracking within and around the fibres that significantly adds to the energy needed to drive the crack. This is illustrated by recent work of A.R. Sanadi, S.V. Prasad and P.K. Rohatgi (*J. Mater. Sci. Lett.* **5**, 395; 1986), who looked at fracture of polyester resins reinforced by fibres of sunhemp. They observed extensive splitting and deformation of the fibres during fracture of the composite, and argue that this would give higher toughness than found in an equivalent composite reinforced with a brittle fibre such as glass. The sunhemp composite has 15 times the work of fracture of the polyester resin alone.

At a recent meeting on ceramics in Pennsylvania, United States (15–19 June 1986), A.G. Evans (University of California, Berkeley) summarized the requirements for the successful application of ceramic (for example, alumina or silicon nitride) parts in engines. These were a toughness of 20–25 MPa m^{1/2}, a strength of 300 MPa and a creep strain of less than 1 per cent at high temperature (>1,000°C). All these properties can be achieved, but not all in the same material at once and not reliably. The toughness of the ceramic can be raised by two methods: one is to incorporate some tetragonal zirconia which raises the toughness by a martensitic transformation process similar to that in steel from 3–5 to 10–20 MPa m^{1/2}. This has the disadvantage that the toughening is not stable to high-temperature treatments. The second is to incorporate fibres or whiskers of another ceramic to make a ceramic-toughened ceramic composite. Because we want to make moulded parts

rather than bars or sheets, the fibres must be randomly oriented to produce a structure very similar to that of the macadamia shell. This structure is quite different from conventional composites where the fibres are either all parallel or are arranged in layers and where a hard, strong fibre is used in a soft but tough matrix, as in glass-fibre reinforced plastics.

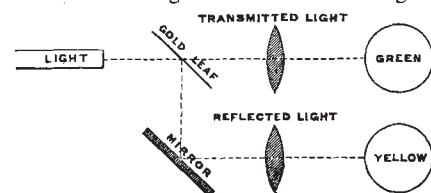
Silicon carbide whiskers about 0.6 µm long are commercially available from pyrolysed rice hulls. These whiskers have been used to reinforce alumina, producing a doubling in toughness to 6–9 MPa m^{1/2}. Because the system was hot-pressed, the fibres were in a random-in-plane orientation rather than isotropic. At loadings higher than 20 vol per cent, the whiskers did not increase toughness because they cause extensive cracking in the alumina during sintering (Wei, G.C. and Becher, P.F. *Am. Ceram. Soc. Bull.* **64**, 298; 1985). The questions that we now wish to ask the nut are: what is the maximum toughening we can expect from whiskers in hard matrix systems; and how can we achieve high loadings of fibres into an isotropic, packed-bundle arrangement? □

Paul Calvert is in the School of Chemistry and Molecular Sciences, University of Sussex, Brighton BN1 9QJ, UK.



100 years ago

WE are apt to think of gold as being essentially and distinctively golden-yellow; it may, however, possess a wide range of colours. A leaf of gold when seen by transmitted light is either green or blue, according to its thickness. Here is such a leaf of green gold as seen when light is actually sent through it, so as to project a green disk on screen. A portion of the light will be reflected from its surface, and this reflected ray may be caught in a mirror and thrown on the screen so that you have, shown side by side, the green disk of transmitted light and the golden one of reflected light from the same leaf of gold.



It would be easy to show that light is similarly affected by other metals; but I have selected gold for the purpose of illustration because it is easy to maintain it in a state of purity, however finely divided it may be.

We must therefore modify any views we may have formed as to a metal having exclusively a special colour of its own, because it will be evident that a particular colour is only due to a definite state of arrangement of its particles.

From *Nature* **35**, 109; 2 December 1886.

Semiconductors

An optical Stark effect observed in gallium arsenide

John Ryan

ELECTROMAGNETIC radiation interacts with atoms to modify many of their fundamental properties including the atomic energy levels themselves. The most famous example of this is the Lamb effect, which occurs spontaneously in the absence of an applied electromagnetic field (in other words the field is 'virtual', arising from quantum fluctuations in the vacuum); in atomic hydrogen small energy-level shifts of the order of 1 GHz arise from the effect. In condensed matter, the intrinsic energy levels are broadened considerably by interatomic interactions so that such small shifts are not normally detectable. However, André Mysyrowicz and colleagues at École Polytechnique in Paris now report (*Phys. Rev. Lett.* **56**, 2741; 1986) the observation of a remarkably large shift of around 500 GHz in the energy levels of the semiconductor gallium arsenide, GaAs. On the basis of their results, the authors believe that the effect can be observed quite generally in semiconductors, and that it has potentially important applications in high-speed optical switching.

The semiconductor studied by the Paris group is a composite material known as a quantum-well structure: it consists of thin slabs of GaAs containing about 50 atomic layers sandwiched between equally thin slabs of GaAlAs such that the potential barrier formed at the interface between the two materials causes electrons and holes to be confined to the GaAs. The energy state under investigation was the intrinsic exciton, which is a hydrogen-like bound state of an electron and hole having a Bohr radius of approximately 100 Å. In GaAs quantum wells, the resonant generation of excitons by light gives rise to a well-defined absorption line which is sufficiently sharp to be a useful probe of electric-field-induced energy shifts.

In the experiment, an electromagnetic field was applied in the form of an intense ultra-short laser pulse of 150 femtosecond duration. The electric field produced is equivalent to a potential difference of about a tenth of a volt over the exciton, which is in fact greater than the electrostatic field experienced by the electron or hole due to the presence of the other. Under these conditions large energy shifts are expected.

Previous experiments by the authors (Hulin, D. *et al.* *Phys. Rev.* **B33**, 4389; 1986) and by a group at AT&T Bell Laboratories (Schmitt-Rink, S., Chemla,

D.S. and Miller, D.A.B. *Phys. Rev.* **B32**, 8027; 1985) had previously detected a pronounced blueshift of the exciton absorption line under conditions of resonant pumping, that is when the laser wavelength coincides with the absorption line. This effect was shown to be due to exciton-exciton interactions, with the shift persisting during the exciton lifetime, typically several nanoseconds. In its latest experiment, the Paris group now detects a large blueshift for non-resonant below-band-gap pumping (where no excitons are generated) which persists only during the pump laser pulse. Mysyrowicz interprets this as evidence of an electric-field-induced energy shift or optical Stark effect.

To explain the result, the authors adopt a quantum-electrodynamical model of 'dressed' excitons which are mixed states

of excitons and photons. The state with n_e excitons and n_p photons $|n_e, n_p\rangle$ is nearly degenerate with the state $|n_e+1, n_p-1\rangle$, and when the exciton-radiation interaction is included, the new mixed states are found to be split in energy by an amount which depends on the strength of the field. This model therefore explains qualitatively the observed blueshift of the exciton absorption line; the magnitude of the shift can also be predicted theoretically, but there is some uncertainty in the actual value of the electric field strength inside the semiconductor that prevents a precise quantitative comparison.

The Paris group has already extended its measurements to bulk GaAs and again detected an optical Stark effect. These are important fundamental observations that open the way to the study of quantum-electrodynamical effects in condensed matter. For the time being, however, the suggestion of practical applications looks somewhat premature because of the very large field strengths required. □

John Ryan is at the Clarendon Laboratory, Oxford OX1 3PU, UK.

Plant pathology

Host antifungal agents

Richard A. Dixon

UNLIKE animals, which can defend themselves against microbial attack by producing circulating antibody proteins, plants respond to fungal and bacterial pathogens by synthesizing a number of proteins that produce a localized 'antimicrobial environment' around the invading pathogen^{1,2}. Lectins are thought to play a role as they have the potential to bind fungal or bacterial cell walls³, and hydrolases (chitinase and 1,3- β -D-glucanase) have the potential to degrade macromolecular components of fungal walls. The report by Schlumbaum *et al.*⁴ on page 365 of this issue now suggests that chitinase is an important component of the plant's antifungal arsenal and shows that the antifungal activity of plant lectins results from contamination with chitinase.

The microbially induced plant gene products that have received most attention have been the enzymes involved in the synthesis of low molecular mass antifungal products termed phytoalexins^{1,2}. A large body of mainly correlative evidence suggests that the *de novo* synthesis of phytoalexins from intermediary metabolites, a process which in legumes such as bean probably requires the induced transcription of about 10 different genes, is a major determinant of active defence against microbial pathogens. This complexity is daunting for the development of disease-

resistant plants using gene-transfer techniques, hence the attraction of chitinase which is a single antifungal product.

Previous reports have shown that this enzyme has no apparent endogenous substrate in plants but can release chitin oligomers from purified fungal cell walls, and is induced in plant tissues by fungal infection or exposure to the gaseous plant growth regulator ethylene. Schlumbaum *et al.*⁴ now show that purified bean chitinase is highly inhibitory to the growth of the fungus *Trichoderma viride*, at concentrations similar to those found in crude extracts from infected or ethylene-treated bean leaves. The antifungal activity of bean leaf extracts is almost completely inhibited by anti-chitinase antibody, suggesting that chitinase is the major protein inhibiting the growth of *Trichoderma* that is present in these extracts. At similar concentrations, chitinase shows the same antifungal activity as the bean phytoalexins but, significantly, it is about 300 times more potent than the lectin wheat-germ agglutinin, whose inhibitory effect on the growth of *Trichoderma* had previously led to the suggestion by Mirelman *et al.*⁵ that plant lectins (particularly those with binding specificity for *N,N'*-diacetylchitobiose and *N,N',N''*-triacylchitotriose) may act as natural antifungal agents.

Schlumbaum *et al.*⁴ have re-evaluated

the antifungal activity of potential chitin-binding lectins. Their results indicate that wheat-germ agglutinin, when prepared by the method of Mirelman *et al.*⁵, is contaminated with chitinase and has antifungal activity. Four commercial preparations of wheat-germ agglutinin, which were not contaminated in this way, could not inhibit the growth of *T. viride*. Unfortunately, wheat chitinase is not recognized by antibodies raised against the bean enzyme, and it was not therefore possible in this case fully to exclude the involvement of other minor components (possibly even other lectins) in the antifungal activity of the preparation. The demonstration of chitinase contamination in antifungal lectin preparations from tomato, potato and pokeweed; its absence from all lectins not showing antifungal activity; and the total inhibition of the antifungal activity of the tomato and potato lectins by anti-bean chitinase antibody strongly suggests that lectins do not inhibit fungal growth. But they still have the potential to function in plant-microbe recognition associated with, for example, the binding of symbiotic bacteria to legume roots or the reception of polysaccharide elicitor molecules whose binding may trigger responses which include the synthesis of phytoalexins and chitinase.

The studies of Schlumbaum *et al.*⁴ have relied upon the non-pathogenic *T. viride* as the test organism; it is therefore now important to consider the effects of bean chitinase on true bean pathogens such as rust (*Uromyces phaseoli*) and anthracnose (*Colletotrichum lindemuthianum*). Induced resistance to the latter organism comprises, in addition to induction of chitinase, the accumulation of several isoflavonoid phytoalexins, hydroxyproline-rich glycoproteins and cell-wall-bound phenolic compounds¹, and the relative importance of these individual components to the overall outcome of the host-pathogen interaction is as yet unknown.

A simple, direct test for the antifungal contribution of chitinase *in vivo* is not obvious; although the enzyme can be induced by ethylene, the relation between ethylene production and pathogen-induced defence responses is still unclear, and the use of inducers and inhibitors of

ethylene production has nearly always yielded equivocal results⁶.

Chitinase complementary DNAs and genes have recently been cloned from both plant (R. Broglie, unpublished; C.J. Lamb, personal communication) and microbial⁷ sources, and it may therefore soon be possible to assess the effects of heterologous chitinases on the disease resistance of transgenic plants. It should be remembered, however, that chitinase alone may not have sufficient activity; Schlumbaum *et al.*⁴ refer in their paper to recent unpublished work suggesting that mixtures of chitinase and 1,3- β -glucanase (which are often co-induced in plants) may be more effective.

The induction of chitinase in infected bean plants may be a direct response to the action of elicitor molecules. A polysaccharide elicitor from the cell walls of *C. lindemuthianum* can induce the enzyme in cultured bean cells⁸. It has been known for some time, however, that chitosan, a par-

tially depolymerized, de-acetylated form of chitin, acts as an elicitor of phytoalexin synthesis in peas and can also inhibit the growth of the pea pathogen *Fusarium solani*⁹. The rapid accumulation of chitosan in infected pea pods would clearly require the action of constitutive chitinase and *N*-acetylglucosaminidase. More recently, chitosan and chitin oligomers have themselves been implicated as elicitors of defence-related lignification in wheat¹⁰. Future work will tell whether more than one chitinase gene exists in some plants, or whether the different functions of chitinase are exclusive to different host-pathogen interactions. In the meantime, the potential importance of a naturally occurring, inducible antifungal agent arising as a single gene product should not be overlooked. □

Richard A. Dixon is in the Department of Biochemistry at Royal Holloway and Bedford New College, University of London, Egham, Surrey, TW20 0EX, UK.

Retroviral immunization

Problems with AIDS vaccines

Peter Newmark

CURRENT hopes for a vaccine against AIDS (acquired immune deficiency syndrome) focus on the use of the whole or part of the envelope protein of the human immunodeficiency virus (HIV) as the protective antigen. It is too soon to say whether or not this approach will succeed but the signs were not encouraging at a meeting held in Paris at the end of October*. Participants were left doubting that envelope protein alone can induce the antibodies that would provide protection against the virus, suspecting that cell-mediated as well as antibody-mediated immunity will be needed, and worried about the availability of an adequate animal model and sufficient animals on which to test potential vaccines.

At present, chimpanzees are the only animals considered to be suitable for testing AIDS vaccines. They do not develop an immune deficiency as a result of exposure to HIV-1 but at least there is a persistent viral infection which should be prevented by any effective vaccine. But first, any potential vaccine needs to be tested in chimpanzees for its ability to stimulate the immune system sufficiently for there to be a good prospect of protection — yet there are insufficient chimpanzees for all the testing that is and will be necessary. The solution, according to Donald Francis (State of California Health and Welfare Agency) is for the initial tests to be carried out in humans.

His suggestion is that the immune system's response to any potential vaccine is first tested on human volunteers, who might particularly be found among health-care workers and monogamous homosexuals. The vaccines inducing the strongest response would then be tested in chimpanzees for their ability to protect the animals against infection with the virus. Success at that state would warrant vaccination trials in humans that fall into 'at-risk' groups.

Meanwhile, at least one very small-scale trial of a potential vaccine in chimpanzees is nearing completion (Shiu-Lok Hu, Oncogen, Seattle). Two animals were immunized with a recombinant vaccinia virus carrying the HIV-1 envelope protein gene. Not only did the animals raise antibodies against the protein but there was also evidence of cell-mediated immunity in the form of proliferating T cells and cytotoxic T cells (which released radio-labelled chromium from target cells infected with the recombinant vaccinia virus. The animals have now been challenged with the AIDS virus and the outcome is awaited with interest.

If this trial fails — and nobody was betting on a successful outcome — it will be particularly bad news in view of the observed immune response, which is the first to include the generation of cytotoxic T cells. It seems unlikely that antibodies alone will provide an effective measure against viral infection and yet most potential vaccines other than the recombinant vaccinia virus induce antibodies but no clear cell-mediated response. Even the antibodies

1. Dixon, R.A. *Biol. Rev.* **61**, 239–291 (1986).
2. Halverson, L.J. & Stacey, G. *Microbiol. Rev.* **50**, 193–255 (1986).
3. Etzler, M.E. *A. Rev. Pl. Physiol.* **36**, 209–234 (1985).
4. Schlumbaum, A., Mauch, F., Vogeli, U. & Boller, T. *Nature* **324**, 365–367 (1986).
5. Mirelman, D., Galun, E., Sharon, N. & Lotan, R. *Nature* **256**, 414–416 (1975).
6. Chappell, J., Hahlbrock, K. & Boller, T. *Planta* **161**, 475–480 (1985).
7. Taylor, J.D., Müller, G., Bond-Nutter, D. & Jones, J.D.G. *Hoppe-Seyler's Z. Biol. Chem.* **367**, (suppl.) 148 (1986).
8. Dixon, R.A. *et al. Phil. Trans. R. Soc.* (in the press).
9. Hadwiger, L.A., Beckmann, J.M. & Adams, M.J. *Pl. Physiol.* **67**, 170–175 (1981).
10. Ride, J.P. *Hoppe-Seyler's Z. Biol. Chem.* **367**, (suppl.) 75 (1986).

*Retroviruses of Human AIDS and Related Animal Diseases, Marnes-La-Coquette, France 29–30 October 1986.

they induce are of limited value in that they are usually unable to neutralize strains of the virus that are not closely related to the immunizing strain (Robin Weiss, Institute of Cancer Research, London; Dani Bolognesi, Duke University Medical Center). To mimic the more broadly neutralizing antibodies present in AIDS patients, it will probably be necessary to use a multivalent vaccine, incorporating the envelope proteins of several different strains of HIV-1 and perhaps also viral proteins that are not exposed on the surface of infected cells and yet might be important for the generation of a cell-mediated response (as in the case of influenza virus — see, for example, Ronald Germain's *News and Views* article in *Nature* 322, 687; 1986).

There will be additional problems in designing a vaccine that provides protection not just against HIV-1 but also against HIV-2 (otherwise known as LAV-2 or HTLV-4). This virus, which is only rather distantly related to HIV-1, is the predominant HIV in western Africa, where it has been found both in association with AIDS (Luc Montagnier, Pasteur Institute) and in many symptom-free people (Myron Essex, Harvard School of Public Health). Essex reported 25 cases where both HIV-1 and HIV-2 were present as well as evidence that cells first infected with HIV-2 can still be infected with HIV-1, albeit with somewhat reduced cytotoxic effects. Complementary data came from Patricia Fultz (Centers for Disease Control, Atlanta), who reported that a single chimpanzee infected with HIV-1 was still susceptible to HIV-2.

Because HIV-2 seems to be closely related to simian immunodeficiency viruses (SIVs), there is some hope that it will be much easier to have an effective animal model for vaccines against AIDS produced by HIV-2 than by HIV-1. Ideally, HIV-2 would be found to produce AIDS in rhesus monkeys or some other species whose availability is not a problem. But first results are not promising: no virus could be isolated from rhesus monkeys following their infection with HIV-2 (Fultz; Ronald Desrosiers, New England Regional Primate Research Center). In the case of two baboons infected with HIV-2, virus was recovered for the first 3–4 weeks but not in the next 8 weeks (Desrosiers).

An alternative approach is to use a SIV as a model for HIV. Indeed, Murray Gardner (University of California, Davis) reported that a killed vaccine was successful in protecting rhesus monkeys against simian AIDS. But the virus that causes the disease is a type D retrovirus and, as such, is unrelated to the other human and simian viruses that cause AIDS, all of which are type C retroviruses. Thus the general relevance of this success is in doubt (see Jerome Groop-

DNA structure

A simple solution to the stability of the double helix?

Maxim Frank-Kamenetskii

WHY are different forms of DNA stable under different ambient conditions? This question becomes more and more important as the structural polymorphism of DNA (the left-handed Z form as well as the right-handed A and B forms) and its biological significance is more generally recognized. But is there a simple answer? Wolfram Saenger *et al.* elsewhere in this issue (*Nature* 324, 385–388; 1986) claim that there is.

Their answer is the principle of hydration economy. In the B form of DNA the phosphate groups are far apart and are independently hydrated. In the A and Z forms the phosphate groups are brought closer together and are bridged by water molecules. Thus, in contrast to the B form, the A and Z forms are economically hydrated. As a result the latter two forms are stable under a water deficit, whereas the former is stable under a water excess.

This simple answer is based on much work in recent years in the field of DNA X-ray crystallography. Until recently X-ray diffraction studies of DNA were restricted to DNA fibres. In those studies water molecules were completely obscured. The structure of the DNA hydration shell emerged only after X-ray crystallography was applied to DNA fragments by A. Rich, R. Dickerson, O. Kennard and their co-workers in the 1980s. The first concept based on these data was the water spine, which is situated in the minor groove of B-DNA and stabilizes its canonical structure (Grew, H.R. and Dickerson, R.E. *J. molec. Biol.* 151, 535–556; 1981). Now Saenger *et al.* introduce the concept of hydration economy.

This novel concept is based on the fact that in the A and Z forms the free oxygen atoms of phosphate groups are, respectively, 5.3 and 5.1 Å apart and are bridged by water molecules. In B-DNA the distance between the oxygen atoms is 6.6 Å, too large for a water molecule to bridge. Hence the major conclusion of Saenger *et al.* is that the A and Z forms should become more advantageous compared with the B form under a water deficit.

This concept readily explains the long-known fact that DNA dehydration (through drying or substituting alcohol for water) induces the B→A transition. In poly(GC)·poly(GC), dehydration induces the B→Z transition. Saenger *et al.* hypothesize that the mechanism of the famous salt-induced B→Z transition in poly(GC)·poly(GC) is the same, as the salt pumps water molecules out of the polynucleotide.

Not everybody will share the latter viewpoint. Many believe that the high-salt B→Z transition is governed by electrostatic interactions of phosphate groups with each other and with the salt ions. In such a polyelectrolyte model, water does not play a structural role and only dampens the interactions between charges. Doubtless the principle of hydration economy is just another factor in a very complex balance of interactions that governs the stability of double helical conformations. But its importance can no longer be ignored. □

Maxim Frank-Kamenetskii is in the Institute of Molecular Genetics at the Academy of Sciences of USSR, Moscow 123182, USSR.

man's *News and Views* article in *Nature* 323, 489; 1986).

Indeed, an initial equivalent trial with a type C SIV was not successful (Desrosiers). In that trial two groups of four rhesus monkeys were vaccinated with an inactivated form of the type C SIV isolated from rhesus monkeys. For one group the virus was given with threonyl muramyl dipeptide (a promising new adjuvant which was used in Gardner's successful vaccine); for the other group the vaccine consisted of viral proteins in the form of an iscom (immune stimulatory complexes in which the proteins are attached to a matrix). Eighteen weeks after vaccination the monkeys were challenged with the SIV. Despite the fact that both vaccines induced a reasonable antibody response (although it was not determined whether

neutralizing antibodies were present), when the monkeys were challenged with the rhesus SIV, the virus became established in the animals with a viraemia that has persisted for fifteen weeks, which is the duration of the experiment so far.

For now, anybody who states that an AIDS vaccine will be available in 3 or 5 or 10 years time is simply hazarding a guess. But reports of at least partly successful vaccines against two type C retroviruses — a recombinant vaccinia vaccine against Friend murine leukaemia virus (Bernard Moss, National Institutes of Health) and an iscom vaccine against feline leukaemia virus (Bror Morein, Swedish University of Agricultural Sciences, Uppsala) — suggest that the problem can be solved. □

Peter Newmark is deputy editor of *Nature*.

Celestial mechanics

Orbital evolution of comets

Julia Heisler

A FLAW in our understanding of the orbital evolution of comets is that the number of short-period comets — those with orbital periods less than 200 years, such as comet Halley — is much greater than theory predicts¹. The discrepancy is enormous; the observed number is two orders of magnitude larger than expected². Relying partly on some newly recognized perturbations acting on comets, Bailey³, on page 350 of this issue, presents a possible solution to this problem.

In 1950 Oort proposed that the Solar System is surrounded by a spherical cloud containing 10^{11} comets; today it is believed that a typical 'Oort cloud' comet orbits the Sun at a distance of 25,000 astronomical units (AU). Until recently it was thought that the orbits of comets are mainly perturbed by random stars that pass within about one parsec of the Solar System. In the past year, a second, more important perturbation source has been recognized. The relatively smooth gravitational potential of the galactic disk in which the Solar System is embedded is responsible for a tidal effect upon comets that is proportional to the total density of material (stars, gas and dark matter) in the solar neighbourhood. (This is analogous to the orbital precession of the Moon because of the pull of the Sun.)

Both perturbations — from passing stars and galactic tide — gradually alter the orbital angular momentum carrying some comets into the inner Solar System and under the gravitational influence of the outer planets. Those comets that pass within 5 AU of the Sun are observable from the Earth and are called 'new' comets, as they are probably making their first passage through the planetary system. A new comet is recognized by its extremely small orbital energy (in absolute value); after one passage through the inner Solar System the orbital energy is typically increased sixfold. Oort cloud comets that pass within 10 AU of the Sun are either ejected from the Solar System entirely or captured into short-period orbits. However, even using both perturbations, the number of short-period comets derived from this capture process is still a small fraction of the observed number.

A solution to this problem is that the short-period comets come from a source other than the Oort cloud, from a second, closer comet reservoir. Current ideas⁴ are that a comet belt extends outwards from the planetary system and departs from the ecliptic to blend in with the spherical outer Oort cloud at about 10,000 AU. Comets at distances between a hundred and a few

thousand AU are virtually unperturbed by stars, galactic tides and planets. The comets from the inner region drift into the planetary system with perihelia well outside the detectable limit of 5 AU, and thus no significant flux of observable new comets is produced. (In the case of the outer Oort cloud, the perturbations are stronger and some comets can be directly perturbed into orbits that are observable from the Earth.) As a result of the lack of observable new comets from the inner Oort cloud, the spatial (or energy) distribution is very uncertain, and *ad hoc* assumptions are unavoidable. However, Bailey¹ turns the short-period comet problem around, and demonstrates that the number of short-period comets constrains the choices, allowing one to give a crude determination of the spatial distribution.

Bailey assumes a spherically symmetric cloud with a power-law energy distribution of unknown index. He demonstrates that up to about 30,000 AU the galactic tide perturbs far more comets into the planetary region than stellar encounters do, and hence one can neglect stellar perturbations when calculating the injection rate into the planetary system. This conclusion has been checked independently by numerical experiments⁵. Bailey computes the fraction that will evolve to short-period orbits and multiplies that fraction by the average lifetime of observed comets to derive the expected total number of short-period comets. To explain both the numbers of new and short-period comets, the energy distribution must be fairly steep. For a comet lifetime of 3,000 years and an inner Oort cloud radius of 3,000 AU, 97 per cent of the comet population must be inside 10,000 AU. This conclusion implies that most short-period comets originate in orbits near the cloud's inner boundary, whereas new comets originate at distances greater than 10,000 AU. The short-period comets come from the inner Oort cloud and the new comets from the outer Oort cloud. Furthermore, using entirely analytical methods and without invoking any other perturbations aside from the galactic tide, Bailey is able to derive the correct number of observed short-period comets.

A massive inner cloud has also been invoked to explain biological extinction events on the Earth as a result of comet showers. The related suggestion that large impact craters and extinctions over the past 250 million years occur periodically every 26–30 million years has received much attention, but there are doubts that the claimed periodicities in the data are

genuine^{6,7}. It is also unlikely that comet showers are responsible for either periodic or random events on such short timescales. Periodic comet shower theories include the proposals of a solar companion (Nemesis)^{8,9}, a tenth planet (Planet X)¹⁰ and catastrophic encounters with giant molecular clouds during galactic plane crossings¹¹. The difficulty with the galactic plane crossing suggestion is that although the Sun does pass through the galactic midplane every 30 million years, the amplitude of its oscillation is small, about 70 parsecs. Giant molecular clouds have a similar scale height, and therefore encounters can occur at all times during the Sun's oscillation, giving no statistically significant periodicity¹³. In fact, despite the Sun's motion, the density of material near the Sun is fairly constant, implying that the comet flux induced by the galactic tide is also fairly uniform. Finally, even random comet showers caused by close stellar encounters are unlikely to explain the crater record because an encounter capable of creating a shower is expected only once every 100 million years⁵. Encounters with giant molecular clouds occur even less frequently, about once every 500 million years. Hence, the crater data may reflect only a few comet showers. This conclusion is supported by identifications of impact material at eight large craters up to 200 million years old; only two of the craters could have been formed by comet impacts, the rest were probably made by asteroids¹⁴.

If Bailey's bold extrapolation of the total comet population is correct, the short-period comet problem is solved; it is satisfying that this problem is soluble using perturbation effects we already understand. However, our theories about the inner region of the Oort cloud are still speculative, and we must proceed carefully, because unlike the situation for the classical Oort cloud, there are no observable new comets that directly sample the proposed inner comet cloud. □

1. Joss, P. C. *Astr. Astrophys.* **25**, 271–273 (1973).
2. Fernández, J. A. & Jockers, K. *Rep. Prog. Phys.* **46**, 665–772 (1983).
3. Bailey, M. E. *Nature* **324**, 350–352 (1986).
4. Hills, J. G., *Astr. J.* **86**, 1730–1740 (1981).
5. Heisler, J., Tremaine, S. & Alcock, C. *The Frequency and Intensity of Comet Showers from the Oort Cloud* preprint (1986).
6. Hoffman, A. *Nature* **315**, 659–662 (1985).
7. Tremaine, S. *The Galaxy and the Solar System*, (eds Bahcall, J. & Smoluchowski, R.) (University of Arizona Press, Tucson 1986).
8. Davis, M., Hut, P. & Muller, R. A. *Nature* **308**, 715–717 (1984).
9. Whitmire, D. P. & Jackson, A. A. Jr *Nature* **308**, 713–715 (1984).
10. Whitmire, D. P. & Matese, J. J. *Nature* **313**, 36–38 (1985).
11. Rampino, M. R. & Stothers, R. B. *Nature* **308**, 709–712 (1984).
12. Weinberg, M. D., Shapiro, S. L. & Wasserman, I. *The Dynamical Fate of Wide Binaries in the Solar Neighborhood* preprint (1985).
13. Thaddeus, P. & Chanan, G. A. *Nature* **314**, 73–75 (1985).
14. Weissman, P. *Nature* **314**, 517–518 (1985).

Julia Heisler is at the Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA.

Role of HTLV III/LAV envelope protein

SIR—Sodrosky *et al.*¹ recently reported that the envelope protein of HTLV III/LAV, the virus that causes AIDS (acquired immune deficiency syndrome), might cause the lysis not only of infected T-helper cells (T-4), but also of uninfected T-4 cells. This would explain the total depletion of T-4 cells, even though only some of them are infected with the virus. I would like to draw attention to another possible role for the protein in the pathophysiology of AIDS.

Developmental stages of AIDS are characterized by a selective major histocompatibility complex II (MHC II) restricted immunodeficiency which shows a striking resemblance to class I graft-versus-host immunodeficiency in mice^{2,3}. This immunodeficiency is immunologically mediated by parental T-4 cell recognition of class II MHC alloantigens on the host's B-lymphocytes and results in a polyclonal proliferation of these B-cells and the formation of autoantibodies against a variety of polymeric antigens and also against envelope proteins of murine leukaemia virus bound on kidney cells⁴. These autoimmune processes lead to different pathophysiological symptoms, such as the selective immune deficit and autoimmune glomerulonephritis⁴.

In the case of AIDS these antibodies might be especially focused on the envelope protein of HTLV III/LAV on T-4 cells. The antibodies against the envelope protein might initiate an autoimmune destruction of T-4 cells so that a vicious circle between T-4-cell immunodeficiency and T-4 directed autoimmunity can develop. This autoimmune destruction of T-helper cells might synergistically amplify the direct T-helper cell impairment by LAV/HTLV III and its envelope protein.

It is noteworthy that other autoimmune phenomena in patients with AIDS also show a striking resemblance to those in class II graft-versus-host disease in mice⁴ and some human connective tissue diseases⁵.

I suspect that not only classical cellular and antibody-mediated cytotoxicity⁶ but also polyclonally activated transformed B cells might play a major role in the pathophysiology of AIDS. Strategies which inhibit this polyclonal B-cell proliferation might therefore delay T-4 deletion in early AIDS.

UWE BICKER

Boehringer Mannheim GmbH,
Sandhofer Strasse 116,
6800 Mannheim 31, FRG

1. Sodrosky J. *et al.* *Nature* **322**, 470-474 (1986).
2. Shearer, G.M. & Moser, M. *Immun. Today* **7**, 34-36 (1986).
3. Moser, M., Iwasaki, T. & Shearer, G.M. *Immun. Rev.* **88**, 135-151 (1985).
4. Gleichmann, E. *Immun. Today* **5**, 324-332 (1984).
5. Lane, H.C. *et al.* *New Engl. J. Med.* **309**, 453-458 (1983).
6. Klatzmann, D. & Montagnier, L. *Nature* **319**, 10-11 (1986).

What is an extended probability?

SIR—While grateful for your open-minded discussion¹ of my review of extended probabilities², I should correct one minor point. I introduced the term 'extended probability' not (only) to avoid the charge of sensation-mongering but because probabilities exceeding unity are closely connected to, and in fact a consequence of, negative probabilities. Hence, the former have to be taken into account — if the latter "exist".

I confess my desire that somebody should give meaning to the term extended probability, but considering that Leonardo of Pisa in the thirteenth century was the first to interpret negative solutions of algebraic equations as debts, and even Descartes in the seventeenth century called them "false solutions" (the common term at that time), I do not believe this will happen in the near future. I have to disagree with the definition of negative probability given by D.E. Parry³. If, for instance, extended probabilities are used to obtain a formal resolution of the Einstein-Podolsky-Rosen paradox⁴ they have to obey the following rule: an event leading with negative probability to a certain re-

sult has to extinguish not only the observation of a (hopefully) preceding exhibition of this result, but has also to extinguish the observation of this event at all, that is it has to reduce the number of accumulated events. (This is why a negative probability cannot be interpreted as a positive probability for observing some extinguishing counter-result, but only as a positive probability for the reduction of the number of preceding events.) And, of course, extended probabilities can only then creep into the picture when we deal with sets of indistinguishable results, as in the domain of quantum mechanics.

Unfortunately, no known physical mechanism allows for this kind of interaction or, therefore, provides a local and realistic resolution of the Einstein-Podolsky-Rosen paradox in terms of extended probabilities. Thus, the paradox will presumably continue to bother 'realistic' physicists, perhaps until this species becomes extinct.

W. MÜCKENHEIM

Südring 10,
D-3400 Göttingen, FRG

1. Maddox, J. *Nature* **320**, 481 (1986).
2. Muckenheimer, W. *Phys. Rep.* **166**, 337 (1986).
3. Parry, D.E. *Nature* **321**, 644 (1986).
4. *Lett. Nuovo Chim.* **35**, 300 (1982).

Migration and hominid bipedalism

SIR—Discoveries in palaeoanthropology have taken the story of hominid evolution back to a time just subsequent to the evolution of bipedalism¹. The outstanding evolutionary question now is: what was the selection pressure that produced bipedalism? The current explanation² suggests that bipedalism evolved when the gathering of plants for food became necessary in a drier savannah habitat, and arms were needed to carry food back to a home site. Although this idea is feasible it is not compelling. We suggest a new idea in which bipedalism developed for long distance migration to scavenge from migrating ungulate populations; the only population that could provide sufficient food by scavenging. Bipedalism was a necessary adaptation to exploit this food supply.

In Serengeti, Tanzania, several ungulate species perform long-distance annual migration³. Migration provides an abundant and constant food supply, and hence these populations are large compared to sedentary populations of the same species. Mammal predators and scavengers do not follow the migrations because their young are slow growing and cannot travel with adults. Consequently breeding predators are sedentary. If a scavenger could migrate it would have access to an abundant and constant supply of carcasses (at least 1 carcass per 20 km² per day)³, an order of magnitude greater than in non-

migratory systems. In fact, humans can find enough to eat by following the wildebeest migration on foot⁴. Almost 70 per cent of carcasses are the result of undernutrition⁵, so hominids would not be displacing predators from their kills.

Two vulture species follow the migrating ungulates, and they are considerably more numerous than sedentary species⁵. Vultures, however, have the disadvantage that they cannot easily break through the skin of intact carcasses. Therefore, there is in Africa an unfilled niche for a mammalian scavenger that can follow migrating ungulates; but such a mammal would need to carry its young.

We suggest that the selection pressure for the evolution of bipedalism in early hominids was access to this new rich and constant food supply. Morphs that had developed the ability to walk long distances would increase rapidly in number and displace the less numerous sedentary quadrupedal types dependent on plant gathering.

We envisage the protohominid quadruped as a plant gatherer and occasional scavenger, much like the baboon. For this type to follow migrating ungulates two essential adaptations are required simultaneously. First, members of the type must carry their young efficiently. This requires arms for carrying, and an upright stance. Chimpanzee females carry their young for at least the first year, so do humans, and thus it is reasonable to assume that early hominids also did so. Pro-

longed upright stance for efficient carrying would require adaptations of the hip joint, such as those in *Australopithecus afarensis* onwards⁶. Second, they must travel long distances efficiently. For this, the primitive foot with opposable big toe must change to the modern foot as a propulsive lever, and this change must coincide with changes in the hip.

The alternative hypothesis is that bipedal hominids were plant gatherers in a savannah home range^{2,6}. Plant gathering required hands for carrying food; gradually the hominids' diet changed to include scavenged meat, using tools. This idea is less compelling for two reasons. First, modern baboons and chimpanzees are plant gatherers, and they can live in hot dry savannahs⁷. Chimpanzees do not travel far (usually less than 4 km a day). Hence plant food gathering in dry environments need not lead to either bipedalism or long distance travel. In contrast, efficient long distance travel (10–20 km a day) is necessary to follow the migrations. If arms are used to carry young, bipedalism is the only efficient means⁸. Second, a plant gatherer within a fixed range has few opportunities for finding ungulate carcasses; mortality of nonmigratory ungulates in Serengeti produces one carcass per month in 10 km². Thus selection to use tools would also be very small.

The migration hypothesis suggests that tool use was an adaptation to speed up the butchering of carcasses and avoid competition with other stronger mammal predators. Indeed, tools were used for scavenging 2 million years ago in Serengeti⁹. The opportunities for hominid migrations to evolve were numerous, for savannah Africa was dominated by migration ecosystems⁵.

The two hypotheses are testable and can be distinguished by the timing of changes in the foot. For migration, hip and foot changes are both linked and essential. Plant gathering, however, does not predict simultaneous changes in the foot because long-distance movement is not essential. Unique hominid foot characters have been identified¹⁰; for example the length of the foot is less than 25 per cent of the leg. In future fossil finds, particularly for the period when bipedalism was evolving, it is such characters that will distinguish between the two hypotheses.

We thank the New York Zoological Society and the Canadian NSERC.

A.R.E. SINCLAIR

Department of Zoology,
University of British Columbia,
Vancouver,
British Columbia V6T 1W5, Canada

MARY D. LEAKEY

Box 30239, Nairobi, Kenya

M. NORTON-GRIFFITHS

Ecosystems Ltd,
Box 21791, Nairobi, Kenya

1. Leakey, M.D. & Hay, R.L. *Nature* **278**, 317–323 (1979).
2. Isaac, G.L. in *Early Hominids of Africa*, 219–254 (Duckworth, London, 1978).
3. Sinclair, A.R.E. & Norton-Griffiths, M. *Serengeti: Dynamics of an Ecosystem* (University of Chicago Press, 1979).
4. Schaller, G.B. & Lowther, G.R. *Southwest. J. Anthropol.* **25**, 307–341 (1969).
5. Houston, D.C. in *Serengeti: Dynamics of an Ecosystem*, 263–286 (University of Chicago Press, 1979).
6. Lovejoy, C.O. *Science* **211**, 341–350 (1981).
7. McGrew, W.C. *et al. J. Hum. Evol.* **10**, 1–18 (1981).
8. Fedak, M.A. & Seeherman, H.J. *Nature* **282**, 713–716 (1979).
9. Shipman, P. *Am. Anthropol.* **88**, 27–43 (1986).
10. Lessertisseur, J. & Jouffroy, F.K. in *Primate Functional Morphology and Evolution*, 327–340 (Mouton, The Hague, 1973).

¹³¹I in milk and rain after Chernobyl

SIR—Radioactivity in rain water attributed to the Chernobyl plant was first detected in Japan on 3 May 1986, a week after the accident¹, and included ¹³¹I. We have monitored ¹³¹I both in milk sold in markets and in fresh milk from dairy cows. We found that the concentration of ¹³¹I in milk in Nagoya, Japan at a distance of 8,000 km from Chernobyl was 4–5 times higher than that in rain water, and that both concentrations decreased with approximately the same effective half-life.

Rain water was collected by a 0.305 m² collector between 3 May and 24 June. The rain water with added carrier was concentrated by evaporation. Gamma-ray spectroscopy showed the principal radionuclides to be ¹³¹I, ¹⁰⁶Ru, ¹³⁴Cs and ¹³⁷Cs. The ¹³¹I concentration ranged from 980 pCi l⁻¹ observed on 3 May to 0.019 pCi l⁻¹ on 24 June, decreasing with an effective half-life of 5.2 ± 0.3 days (Fig. 1).

The Ministry of Science and Technology announced that, in milk between 4 May and 12 May, ¹³¹I activity ranged from 57 pCi l⁻¹ to 400 pCi l⁻¹ fresh milk and from 20 pCi l⁻¹ to 160 pCi l⁻¹ for commercial milk. The northern jetstream which may have

carried radioactive materials to Japan, passed over Nagoya, in central Japan. We accordingly began a survey of the ¹³¹I concentration in milk from five local dairies. Because the ¹³¹I concentration varied from dairy to dairy, no doubt due to the variety of feed furnished to cows, we decided to monitor fresh milk from individual cows mainly fed on fresh grass.

Fresh milk sampled from 20 dairy cows at the university farm affiliated to the School of Agriculture of Nagoya University, was monitored between 19 May and 14 July. Sixteen cows were of Holstein breed, and four were first-generation crosses between the Japanese breed and the Aberdeen Angus. The cattle were each fed on fresh Italian ryegrass (34–35 kg per day) and concentrates (5–6 kg per day). The total milk volume from 20 cows amounts to 300–330 litres a day. One litre weekly samples were placed in plastic bottles and mounted on a Ge(Li) semiconductor detector for gamma-ray spectroscopy. ¹³¹I was the only radionuclide identified in almost all of the samples, but a few gave extremely low ¹³⁷Cs signals. For the last two samples, ¹³¹I was chemically extracted and measured with a well-type Ge detector. The ¹³¹I concentration decreased with a half-life of 5.0 ± 0.2 days (Fig. 1), similar to the half-life in dairy milk reported in Britain².

The ¹³¹I concentration in milk was about 4–5 times greater than that in rainwater between 19 May and 14 July, a rainy season in Japan. Judging from the ¹³¹I concentration in rainwater and from that in milk announced by the ministry, it is reasonable to assume that the concentration in fresh milk reached a maximum on 4 May and declined with a half-life of 5 days.

KUNIHIDE NISHIZAWA

KENZO TAKATA

NOBUYOSHI HAMADA

YOSHIMUNE OGATA

SADAO KOJIMA

Radioisotope Centre,

OKITSUGU YAMASHITA

MITSUAKI OHSHIMA

YOSHIMASA KAYAMA

School of Agriculture,

Nagoya University,

Chikusa, Nagoya 464, Japan

1. Aoyama, M. *et al. Nature* **321**, 819–820 (1986).

2. Hill, C.R. *et al. Nature* **321**, 655–666 (1986).

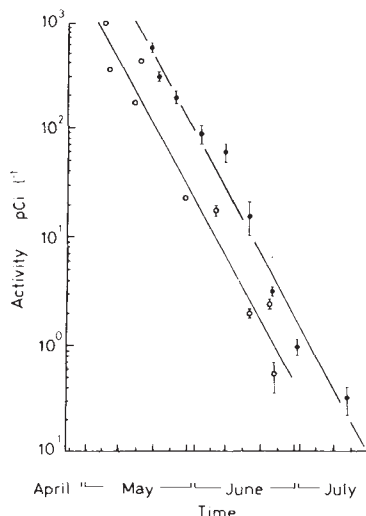


Fig. 1. The ¹³¹I concentration in dairy cow milk (black dots) and rain water (open circles) in Nagoya from 26 April, the date of the Chernobyl accident, to 20 July 1986.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*, but those that do should not be highly technical comments on Articles or Letters (where the Matters Arising section remains appropriate). □

The patient's dilemma

D.J. Weatherall

Unnatural Selection? Coming to Terms with the New Genetics. By Edward Yoxen. William Heinemann: 1986. Pp. 208. £12.95.

GEORGE Bernard Shaw once observed that all professions are conspiracies against the laity, a view that has permeated much of the bad press to which doctors have been subjected in recent years. What has gone wrong? Paradoxically, the successes of the medical sciences in the post-war period, which saw the development of antibiotics and the eradication of many childhood infections by vaccination programmes, may be partly to blame. These remarkable achievements raised the expectation that the control of cancer, heart disease, arthritis and other major killers would soon follow. When it became apparent that this was not to be, and that these conditions could only be managed, often unsatisfactorily, by increasingly sophisticated patch-up procedures involving the use of expensive drugs, a bewildering array of machinery and the replacement of almost every organ except the brain, a sense of disillusionment with scientific medicine set in. Hospitals came to be perceived as terrifying places to visit, let alone in which to be ill, and it was widely believed that the pastoral needs of patients had become lost in a maze of high technology. Alternative forms of therapy were sought, and "holistic" medicine, a term which simply means good doctoring, was rediscovered.

It is hardly surprising, therefore, that when terms such as *in vitro* fertilization and genetic engineering appeared on the scene further public suspicion was generated; the technocrats were at it again, now breeding monsters in test tubes and tinkering with people's genes. The situation was exacerbated by sensational television programmes which featured parents wandering around "gene supermarkets" stocking their baskets with traits that they hoped to see expressed in their children. Memories of Nazi Germany and the eugenics movement were revived, committees of inquiry set up, and hastily constructed and ill-thought-out bills put before the British Parliament in an attempt to prevent any form of research involving human embryos.

Considering the complexities of molecular and cell biology, and the speed with which these fields have advanced over the past few years, it is not surprising that the clinical applications of these new techniques have been greeted with apprehension; the medical profession itself is only just starting to get to grips with their potential. Clearly, a long period of public

debate and reflection is required before we shall be in a position to assess how far we wish to go in interfering with our reproductive activities and manipulating our genomes. This new book by Edward Yoxen should be a valuable catalyst for pointing the discussions in a sensible direction.

Like Jane Austen and J.D. Salinger, Yoxen believes in the analeptic effect of a racy opening paragraph: "There is a great deal of sex in this book, quite enough to shock the faint at heart". After an introductory section (which rapidly disillusiones even the not so faint at heart), successive chapters deal with artificial insemination, *in vitro* fertilization, surrogacy, sex predetermination, antenatal diagnosis, new forms of genetic intervention and gene therapy. In each of these controversial areas the author faces up to the ethical and pastoral problems that are posed and gives a balanced discussion of the benefits and risks involved. By and large he subscribes to the value of these developments, while at the same time showing great sensitivity to the views of those who find many of them distasteful. It is to be hoped that his sane account of such an emotionally-toned topic will be read by teachers, politicians, church leaders and, in particular, members of the many pressure groups that are attempting to prevent further research without evaluating the evidence.

Yoxen makes little attempt to explain the scientific basis of the advances in reproductive physiology and genetics that have given us these powerful new tools, and, in most cases, simply states what is or is likely to be possible and the moral dilemmas that may arise. Presumably he feels that this is as far as he can go at the moment, given the level of biological sophistication in society. Many will agree. In discussing the problem of scientific communication, Francis Crick has described how he had to explain to a lady visitor to the Salk Institute that worms are not spontaneously generated in apples. And, as Yoxen points out, it is not all that long since those of the British aristocracy who wanted male heirs submitted to hemicastration in the belief that sperm from the right testicle produce sons — *noblesse oblige* indeed.

Such tales may not be the best guide to the level of scientific knowledge in the world at large, but there is no doubt that if Yoxen's goal is to be reached — that of an individual who can say "I fully understand the step that I am taking, I have chosen it

freely, having sought advice and considered the options available, and I recognize my responsibilities to myself and to others in what I do here" — we shall have to do some fundamental rethinking about education. To no country does this apply more than Britain, in which children are encouraged to specialize at the age of 14, and in many cases to ignore science completely thereafter, and which, with a few exceptions, is very badly served by its media in scientific matters. Thus far, prejudice and ignorance have characterized many of the views on the topics discussed in this book. Anyone who doubts that there is a problem should read *Hansard's* account of the parliamentary debate on embryo research.

There is another question that is touched on several times in Yoxen's book, but which is never fully developed. What is it that motivates scientists in this field? Why is it, for example, that in the United States there may well be nearly as many research workers trying to replace the defective gene in a rare genetic form of immune deficiency as there are patients with the disorder? And why has at least one worker jumped the gun in a premature attempt at gene therapy? Are patients becoming stepping stones on the way to scientific "firsts" and fame and fortune? As part of the dialogue that is essential if public confidence in modern medicine is to be restored, biologists and medical scientists will have to provide better insights into how science works. They should also ask themselves why it appears that so much talent is being directed into research that gives rise to increasing public concern when, at the same time, there seems to be much less interest in many of the more immediate medical problems of society. □

D.J. Weatherall is Nuffield Professor of Clinical Medicine at the University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

Taking the soft tomato

THE business of reviewing scientific books is one apart from the reviewing of works of literature. But anyone whose book has been roughly handled in a review can find comfort in *Rotten Reviews: A Literary Companion*, which contains such comments as this: "Here all the faults of *Jane Eyre* (by Charlotte Brontë) are magnified a thousand fold, and the only consolation which we have in reflecting upon it is that it will never be widely read". Thus one James Lorimer in a review of Emily Brontë's *Wuthering Heights* (1847).

Rotten Reviews, edited by Bill Henderson, is a 93-page, small-format volume which contains extracts from 175 scathing notices of books that have proved to be of enduring appeal. Publisher is Pushcart Press, and distributor W.W. Norton, New York. Price is \$12.50.

Beastly creations

Martin Kemp

The Rhinoceros from Dürer to Stubbs, 1515-1799. By T. H. Clarke. *Sotheby's Publications/Philip Wilson, Russell Chambers, Covent Garden, London WC2E 8AA, UK: 1986. Pp. 219. £29.50.*

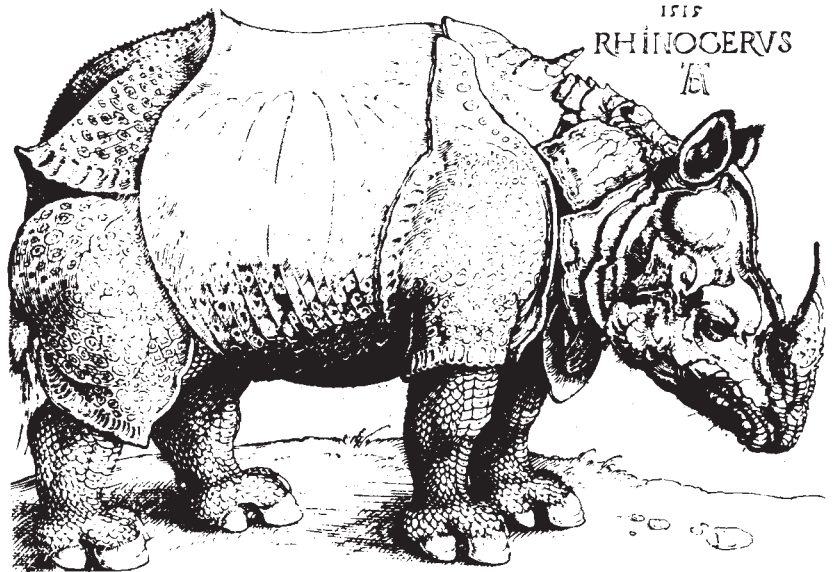
RHINOCEROTICA rules. We might expect an obsessional study of images of a single exotic animal over three centuries to be rich with informative detail. We would not necessarily expect it to be fun. T. H. Clarke's book is both.

The author does not attempt to rival the scholarly detail of L. C. Rookmaaker's *Bibliography of the Rhinoceros* (A.A. Balkema, Rotterdam; 1983). Rather, he has provided a lively compendium of artistic images, ranging from soberly naturalistic records to elaborate fantasies in all kinds of media — painting, sculpture, drawing, engraving, embroideries, tapestry, pottery, porcelain, glass, furniture, and arms and armour. The text is written with a nice wit that only occasionally becomes tiresome.

The rhinoceros is unusually well suited for such a study. It was a rare enough sight for European observers to be the subject of wonder and even fantasy, yet there are enough documented instances of rhinos on European tours to provide a series of identifiable animals from which the great majority of the images sprang. Clarke discusses eight peripatetic visitors between 1515 and the 1790s. In addition to the host of craftsmen who took advantage of the rhinomania occasioned by exhibition of the animals by showmen, a number of accomplished artists were involved. To mention only Albrecht Dürer, Hans Burgkmair, Pietro Longhi, Jean-Baptiste Oudry and George Stubbs will give some idea of the quality and the geographical range.

The involvement of considerable artistic talents in naturalistic illustration has obvious advantages. What is not so readily recognized is that their sheer skill brings dangers as well as benefits. The kind of artistry developed during the Renaissance revolution in representation — and so vividly manifested in Dürer's art — could be used to endow an image of fantasy with a sense of "naturalistic" credibility to match that of a representation taken from life. A skilful artist can make a dragon look as credible as a rhinoceros for someone who has seen neither.

This kind of potency in "fantastic naturalism" explains the immense hold of Dürer's engraving, which, as Clarke indicates, was based on secondary visual and written sources. Dürer's transformation of the leathery rhinoceros into a fearsome piece of late-mediaeval siege machinery



Dürer's rhino, 1515 — "a fearsome piece of late mediaeval siege machinery"

resulted in an image of such power that it remained truer to the emotional response evoked by the animal than the more naturalistic versions. Clarke provides an abundant body of visual evidence to show that naturalistic illustration is a double-edged weapon, though he does not grapple head-on with this fascinating question.

The greatest strength of the book is the remarkable range of media, particularly from the applied arts, that is illustrated and discussed with authority. Its two most obvious weaknesses relate to the history of natural history on one hand, and iconographic art history on the other. Historians of natural history will be

disappointed that the great succession of zoological encyclopaedias is so little exploited, while art historians might reasonably have expected a fuller discussion of the emblematic and symbolic meaning, including the sexual connotations of the "horn". These are not separate matters; for the encyclopaedist, nature was full of meaning, just as meaning for the emblematicist was regularly borne in a natural vehicle. Clarke's approach largely excludes such issues, but it does permit a wide measure of visual delight. □

Martin Kemp is Professor in the Department of Art History, University of St Andrews, Fife KY16 9AL, UK.

Nervous activity

D. Regan

Evoked Potentials. Edited by Roger Q. Cracco and Ivan Bodis-Wollner. Alan R. Liss/Wiley: 1986. Pp. 551. \$96, £70.

OUR knowledge of the cellular and neuropharmacological organization of the brains of experimental animals has advanced enormously over the past decade. How this new knowledge can be used to deepen our understanding of human vision, hearing and motor control, and how it can help to prevent and alleviate visual, hearing and spinal cord disorders, is an immediately challenging problem. It is unfortunate that many of the important advances in neuroscience have been achieved through techniques that, for ethical reasons, can seldom be used in man. A major role of evoked potential recording is to bridge the gap between the human brain and the results from cellular-level experiments on animals.

The term "evoked potential" (EP) covers the electrical responses of the entire peripheral and central nervous system, including peripheral nerve action potentials, spinal cord responses to stimulation of peripheral nerves, the responses of the auditory brainstem, midbrain and cerebral cortex to sounds, the responses of somatosensory brainstem and cerebral cortex to tactile stimulation of the skin, and the responses of the retina and visual cortex to light. The term also covers the electrical activities of the brain associated with the evaluation of the meaning of a stimulus, such as a spoken word, rather than with the physical nature of the stimulus. In this book the "evoked potential" takes on an even wider meaning: Williamson and Kaufman describe how the magnetic responses, 10^{-7} weaker than the Earth's field, that are produced from neurones in the depths of the brain can be clearly picked up by a quantum interference device placed near the head, and the location of the neurones defined to within a few millimetres!

Evoked potentials can be recorded

from metal electrodes held onto the scalp or over the spine without discomfort or risk. Although they are very small (often less than one microvolt), a domestic microcomputer is quite adequate to extract the EP from the electrical "noise" produced by muscle and by the rest of the brain. Evoked potentials thus offer access to the electrical activity of the human brain's visual, auditory and tactile areas while the brain's owner is describing and measuring what he or she sees, hears and feels. In this way the technique can help us to understand the neural activities that underlie normal sensory experience. By the same token, EP recording is a tool for investigating the neural abnormalities associated with disorders of sight, hearing and motor control in human adults, infants and neonates.

This book contains a selective collection of 49 admirably concise mini-reviews on a wide variety of topics, each written by an internationally recognized authority. Illustrating how EPs can be used functionally to "dissect" a sensory pathway, Bodis-Wollner's opening chapter describes the parallel visual subsystems for pattern, flicker, motion and so on. Arguably, the most promising clinical role of the EP technique is in studying disease mechanisms in paediatric patients. After pointing out that the early identification of hearing loss can avoid the disastrous consequences of speech delay, Galambos, and Stockard and Stockard, go on to review the recent advances in identifying hearing defects in newborns by EP recording. Sokol, and Levi and Manny, describe how visual EP recording can aid the management of infants with amblyopic visual loss and help to prevent lasting visual impairment. The possibility of using spinal EPs to assess spinal cord trauma incurred during birth is reviewed by Cracco and Cracco. The applications of EPs in neurology and ophthalmology are extensively discussed, including a useful analysis of visual evoked potentials in multiple sclerosis by Blumhardt and three papers on Parkinson's disease and neurotransmitter abnormalities. The sections on surgical monitoring and on psychiatric disorders in adults and children are instructive and thorough. The book contains an excellent group of papers on the neural sources of sensory and cognitive EPs.

The editors are to be congratulated on putting together such a coherent survey of basic and clinical human electrophysiology. All 49 chapters are carefully and concisely written, and are remarkably free from errors. A uniformly high standard is maintained throughout, and several chapters are quite outstanding. This is the best and most up-to-date book on the topic. □

D. Regan is a Professor in the Departments of Ophthalmology and Medicine, Dalhousie University, Gerard Hall, 5303 Morris Street, Halifax B3J 1B6, Canada.

Developing policies

Ronald Coleman

The Biotechnological Challenge. Edited by S. Jacobsson, A. Jamison and H. Rothman. Cambridge University Press: 1986. Pp. 181. £22.50, \$34.50.

MANY nations, both developed and developing, have attempted to define strategies for the application of biotechnology to improve their overall national economic performance over the next 10 to 20 years. Any such strategy must take account of the science base of the country concerned, the availability of trained manpower and the state of other, complementary technologies, as well as the overall national industrial performance and the ability to deliver goods to the marketplace. This book offers advice on the potential of biotechnology for developing countries.

The editors have brought together experts in their respective fields to outline the current status of the technology, and also present some case-studies relevant to developing countries. For policy-makers, the techniques underpinning modern biotechnology — genetic engineering, and enzymes for biotransformations and fermentation — are well presented in adequate detail. But the choice of case-studies could have been better, and it is unfortunate that none of them relates to health care.

One of the studies deals with the production of high-fructose corn syrup. This is the largest scale industrial use of modern biotechnology, and it has been successful in the developed world, particularly the United States, because the syrup can compete successfully, both in performance and price, with a traditional agricultural product (in this case sugar). But in a developing country sugar production must surely remain the preferred route for the supply of sweeteners: with corn syrup, the possibility of competing successfully for export markets is remote, and it makes no sense to import the necessary technology for a new sweetener in order to achieve a marginal improvement for home markets.

The account of the Brazilian experience in the conversion of sugar cane into alcohol, as a replacement for petroleum, is much more illuminating. Simple economics suggest that the process is less efficient than selling sugar on the world market and using the income to import petroleum. Given the current constraints on Brazil, however, the strategy seems to be both appropriate and successful. The natural advantages of an equatorial country — abundant sunshine, rain and cheap labour — have been used effectively to incorporate a biotechnological process in-

to a large national programme to develop an alternative energy source. Part of the Brazilian experiment is the forward integration of the programme so that the alcohol derived from sugar is used in specially designed engines produced internally, further reducing the outflow of foreign currency. There is little prospect of much competition from manufacturers in the developed world, and the possibility of export opportunities to other countries with a capability of producing alcohol but with a less developed industrial base.

The discussions of biotechnology in food production, particularly aimed at an

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Colour display — the picture (of interferon crystals) is reproduced from Biotechnology: Strategies for Life, by Elizabeth Antébi and David Fishlock, a large-format, highly-illustrated popular book newly published by MIT Press. Price is \$39.95, £29.95.

increase in protein for developing countries, are dated and do not give adequate recognition to the current surpluses of carbohydrates everywhere in the world except Africa. The use of simple technologies, such as the immediate drying of harvested materials to prevent spoilage by fungi, is likely to give a much better return than investment in advanced biotechnological techniques.

It is an especial pity that the book does not include a concluding chapter by the editors, with advice on the prerequisites of a successful process or industry based on the biological sciences. The technology itself is only a small part of the overall picture, and at this stage is probably best imported through licence agreements or joint ventures with leading companies. In the longer term, the commercialization of biotechnology will require substantial investment in trained manpower, and the development of a broad-based research and development programme, and industrial production facilities. Meanwhile, the best hope for the developing countries must be in forming suitable partnerships with other nations. *The Biotechnological Challenge* contains this message, but in only muted tones. □

Ronald Coleman is the Government Chemist in the Department of Trade and Industry, Cornwall House, Waterloo Road, London SE1 8XY, UK.

Distillation to the basics

D.H. Watson

Fundamental Virology. Edited by Bernard N. Fields and David M. Knipe. Raven: 1986. Pp.784. \$54.50.

JENNER may seem to have been anachronistically prescient in writing of smallpox virus. Of course, he only used "virus" in the then-conventional sense of something unpleasant (literally venom or poison). The more restrictive use in the present century was promoted by the observations of Iwanowski and Beijerinck that the causal agent of tobacco mosaic passes through a filter capable of impeding bacteria. Similar properties were soon demonstrated for some organisms infecting animals, followed by recognition of bacterial viruses which, despite their later discovery, were to be the main focus of the early years of virology. A notable landmark was the introduction of physical methods and, in particular, of a rigorous quantitative approach, through the recruitment to the discipline of some distinguished physical scientists (whose scientific emigration has sometimes been attributed to their post-war disenchantment with aspects of modern physics).

Real advance in animal virology only came when animal cells could be cultivated with an efficiency approaching that of their bacterial counterparts. Most importantly, this led to Dulbecco's development in 1982 of reproducible plaque assays for animal viruses. Succeeding years saw accelerating progress in animal virology, which both benefited from and contributed to the growing power of molecular biology.

More recently, it has become clear that the molecular approach to the study of viruses causing animal and, in particular, human disease can lead to rational approaches to chemotherapy, new kinds of vaccines and useful insights into mechanisms of pathogenesis. As a result, modern textbooks on "medical" virology now refer to molecular aspects of the subject, as well as the more traditional clinical aspects of virus infections. The ultimate in this genre must be the encyclopaedic *Virology*, edited by Bernard Fields *et al.*, and published by Raven last year.

Fundamental Virology is a derivative ("exprint" perhaps?) of that earlier volume. It consists, in half the page extent of its parent, of the more basic, rather than the clinical or applied chapters.

The preface concludes with admirable confidence that the text will be useful for courses in general and molecular virology. I agree, but feel the editors may be a little too sanguine in anticipating that it will be

used by senior undergraduates, although it will inevitably be the standard reference book at this level. It should certainly be used and studied at the postgraduate level, and by teachers and researchers.

It is perhaps a little harsh to cavil at the coverage of what is a large volume, even in the magnum, rather than jeroboam, version. Nevertheless, the title is a little misleading. Inevitably, from its provenance, the book deals with the fundamental virology of human viruses, whereas any course on molecular virology must by necessity address bacterial, plant and other animal viruses.

This apart, the volume covers its chosen field comprehensively and all of the chapters have been written by experts on their topics. No other book has such extensive coverage which is so up-to-date, with references up to 1984. □

D.H. Watson is a Professor in the Department of Microbiology, University of Leeds, Leeds LS2 9JT, UK.

Back to the future

John Treherne

Trillion Year Spree: The History of Science Fiction. By Brian W. Aldiss with David Wingrove. Gollancz/Atheneum: 1986. Pp.511. Hbk £15, \$19.95; pbk £9.95.

BRIAN Aldiss and David Wingrove cast a wide net in their 500-page history of science fiction, a revised edition of Aldiss's *Billion Year Spree* published in 1973. Inhabitants of their pantheon range from Gilgamesh and Disraeli to Mickey Mouse and Ray Bradbury. Only poor Fred Hoyle, one of the few scientific professionals among British SF writers, is missing.

It all started in the halcyon years of the Regency, with Victor Frankenstein and his 19-year-old creator, Mary Shelley. Before Mary, there was only what the authors call "ur-scientific fiction" (that is, before the genre originated), with such promising "ur-SF" types as Dante, Voltaire, Daniel Defoe, Will Shakespeare, Erasmus Darwin and the writers of *Genesis*. After Mary, come the real SF-pioneers — Edgar Allen Poe and Bram Stoker, sustained by gothic nightmares of nasty things in dark airless chambers, headless corpses and Count Dracula; followed by Jules Verne, H.G. Wells and Conan Doyle — with giant space rockets, fearful weapons, squidgy creatures in long-legged machines on Cobham Common and the Earth ploughing through toxic interstellar gas — and Edgar Rice Burrows (Tarzan) and William Hope Hodgson, with swine-things that come up from the cellar while the Sun goes dull in the sky.

Meanwhile, "post-ur" auxiliaries were

obligingly scribbling away in the wings: Samuel Butler and William Morris (with their Utopias); Thomas Hardy (whose astronomer-hero, Swinburn St Cleves, gazes at the night sky with Lady Constantine wondering if "horrid monsters lie up there?"); Robert Louis Stevenson (Dr Jekyll, of course); even the gentle, consumptive Richard Jefferies (imagining the worst in *After London*) and many, many more. All are lovingly recruited to the SF cause and accorded lengthy quotation.

Then on past Freud and the Nazi party; Čapek and Kafka, robots and *RUR*, the destruction of civilization; Aldous Huxley and the imbecility of industrial progress; Hollywood and C.S. Lewis; to the whole tottering panoply of late-twentieth-century SF with its intergalactic wars and time warps, bug-eyed monsters (still), supermen and genetic disasters, occasional romps on the wilder shores of psychology and pharmacology and, most surprising of all, the return to pure gothic fantasy with goblins and wizards, absolute evil, sand-worm gods and beautiful princesses in Middle Earth, Dune and Gormenghast.

For most of us, I suspect, encounters with SF are limited to youthful passions for Flash Gordon, Dr Who or Captain Kirk and, in adult years, the occasional Asimov and Aldiss snatched in desperation at Heathrow or from a railway book-stall. The first requirement for the uncommitted SF reader — suspension of disbelief — can be difficult, especially for those of a pedestrian turn of mind. It is not only the convoluted plots and scientific improbability, there are all those silly names. Even Doris Lessing becomes a challenge in this scenario of one of her SF novels:

Two mighty Galactic Empires, one — Canopus — spiritual, one — Sirius — technological, are interested in the degenerated planet Shikasta, once a paradise world called Rohanda. Agents from Canopus and Sirius walk its surface and witness stages in its development and decay...

But for Aldiss and Wingrove this is a serious business:

Science fiction is the search for a definition of mankind and his status in the universe which will stand in our advanced but confused state of knowledge (science), and is characteristically cast in the Gothic or post-Gothic mode.

Undoubtedly they are right when the writers are Wells or Huxley or Kafka and, perhaps, to some extent even when they are L. Ron Hubbard or Arthur C. Clarke. Yet I suspect that most SF readers, like their Edwardian predecessors, are after a rattling good yarn in the realms of ultimate fantasy. And this SF clearly provides in great abundance, as is charted in this readable, scholarly and — despite much good humoured bravado — slightly apologetic book. □

John Treherne, a scientist and novelist, is at Downing College, Cambridge CB2 1DQ, UK. His most recent novel is *Mangrove Chronicles* (Jonathan Cape, 1986).

A renaissance in the making

An OECD survey in 1984 placed Spain and Portugal among nations that "give little or no priority to research and development". But they now seem aware of the need to build scientific resources.

It is hard to square the gloomy judgements of the OECD examiners with the reality of a visit to Iberia. Whether Madrid or Lisbon, science administrators are full of enthusiasm for the future and confident that much can be achieved. It is perhaps after all only human nature: things really are getting better and better and although research investment lags far behind Europe's major nations there is none of the despondency that comes with living in the declining north.

Wider horizons came with the ending of the dictatorships that had governed Spain and Portugal. Although a decade has passed since the first heady days of freedom ("when we talked politics 24 hours a day because we had never been able to before" as one researcher put it) there is still a sense that new opportunities are opening up. This year it is entry into the EEC.

Major challenges come with membership. Both Spain and Portugal are predominantly agricultural nations and lack powerful multinational companies. And while both receive enormous numbers of tourists each year (Spain considerably more than its own population), neither wants to see its involvement with Europe

Names, money and thanks

IBERIAN names, as this correspondent found out, are rather complicated. Having signed my full name, Alun Mark Anderson, in a small hostel in Bilbao I found myself being addressed as "Señor Mark" for the duration of my stay. I might, I am told, have equally well have been addressed as "Señor Mark Anderson" or even "Don Alun" but no one would have committed the foolishness of calling me "Señor Anderson", my presumed maternal surname.

If surnames become muddled in the articles that follow, or if I have given some people more surnames than others, I apologize in advance for my ignorance. Doubtless there will be other errors for this survey is based on a trip to Spain and Portugal that was all too short. Whatever errors there are, they will not be the fault of the people who gave me special help in planning my trip and whose knowledge of Spanish (or Portuguese) science is much deeper than my own.

In Madrid special thanks go to Dr Paul S. Dick, Science Officer at the British Council and his staff for their hospitality and for arranging many of my visits. In Barcelona Dr Pedro Puigdomènech and Professor Juan Subirana deserve equal thanks for looking after every aspect of my visit. And in Lisbon various anonymous members of the Ministry of Education deserve thanks for arranging my schedule.

NOTE: £1 = 198 Spanish pesetas = 229 Portuguese escudos. □

end in economic colonization by the wealthier north.

The increase in agricultural and industrial productivity that both nations now urgently need is inevitably entwined with scientific and technological development. The scale of the problems faced by the two nations is, however, quite different. Spain has four times the population and six times the land area of Portugal. More important, it has a gross domestic product (GDP) per head of £3,087 (1983 figures) that puts it within shouting distance of those, like Britain (GDP per head, £5,542), at the back of the pack of advanced industrial nations. Portugal, with a GDP per head of £1,380, is still at an intermediate stage of development with parts of its agricultural system a century behind the rest of Europe.

Both countries have a history of spending appallingly little on research and development. Even the most optimistic estimates suggest that Spain spends 0.6 per cent of its GDP on research and development and Portugal 0.8 per cent, in comparison with Britain's 2.4 per cent.

Both countries inherited from their dictatorial past institutional structures ill-equipped to deal with rapid change. Their research organizations were fragmented, with different sections of activity separated from one another, and research separated from teaching in the universities. Arbitrary and autocratic control was exerted from above.

The first wave of change was catalysed by young and impatient men and women who had been trained abroad. Even before the dictatorships vanished, many of the brightest students had begun to go to foreign laboratories to gain doctoral degrees. The returnees now provide the core of many of the research groups of international standing and provide the impetus for change.

Spain has moved forward at some speed. Structures have changed and attitudes sometimes seem to be struggling to catch up. In the universities and research institutes, the psychology of the Franco era was expressed by autocratic professors presiding over miniature empires, and in concern with international politics and the skillful currying of favour over the cultivation and reward of excellence. The old Spain is still visible in the laboratories but it is dying, laid low by increasingly open competition for positions and grants.

Overall coordination of the whole spec-

trum of research is now being installed with a new law and a new interministerial body. National priorities will be set and national plans made. A centralized form of administration thus returns, but the government assures doubters that Franco's ghost will not be found within it: researchers have never had greater control over their day-to-day affairs. The new body will, however, have to choose the area in which grants will be given out, while still bearing in mind that the results of much of the research must eventually find application in Spain's agriculture and industry.

That this can be a difficult task is suggested by the failure of a national programme for biomass development launched by the new commission's predecessor. What makes picking winning areas particularly hard is the lack of clear signals from industry of the direction in which it is going. Research spending by industry has always been very low. The new commission will thus have the double task of promoting research that can find application while convincing industry that it should be taken seriously. The fear is that it will be easier to follow international fashions (biotechnology, electronics and new materials are bound to be the new litany) than to find areas in which Spain has a unique chance of success.

The pace of reform in Portugal has been slower. A perfectly good organization to coordinate science and technology exists, but only gradually is it being given the money and powers to do its job.

Meanwhile different research organizations and different sectors of the government happily go their own ways without any real coordination from the top and a curious university research system keeps hundreds of small groups isolated from one another without any real incentive to achieve excellence.

Portugal is saved from its lack of coherent planning by its size. The research community is still small enough for almost everyone to know everyone else, which often means that bureaucratic barriers are not so formidable in the laboratories themselves. Wisely, Portugal is concentrating efforts to develop its backward inner regions through the establishment of new universities and polytechnics with an emphasis on agriculture. Here the scope for improvement is so great that even a little technology transfer is likely to work miracles. □

Spain: political change

A high-speed ride to democracy

IN June of this year the Spanish people re-elected a socialist government with one of the largest majorities in Europe. A little less than ten years earlier that party was illegal, defined itself as "Marxist...rejecting any path of accommodation to capitalism" and had a revolutionary programme intended "to socialize production, distribution and exchange by the working class".

In power the party has taken a very different line; it won a referendum to keep Spain in the North Atlantic Treaty Organization (NATO), took Spain into the EEC and is now busy selling off the vast state holdings accumulated during the Franco era to private industry. But that a radical party transformed itself into something like a conventional European social democratic party is but one of the wholly unexpected twists of history that has brought Spain, in ten years, from a fascist dictatorship to a parliamentary monarchy.

The roots of the peaceful transition can be found in the late 1950s when, with Spain close to bankruptcy, Franco allowed the economy to be liberated. Spain was opened to foreign investment and foreigners and as the tourists flocked in, hundreds of thousands of Spaniards went to work in the then booming industries of northern Europe. The results were dramatic; between 1961 and 1973, the "years of development", Spain grew faster than any country except Japan. By the time the boom was over in 1974, Spain found itself with a society much of which was middle class and which had a vested interest in the continuing stability of the country. In 1975 Franco died and a series of new actors appeared on the political stage.

Six years before his death Franco appointed his own heir (a prerogative he had granted himself). His choice — Juan Carlos, the grandson of the last king of Spain, Alfonso XIII, who been brought up under the tutelage of General Franco — was not a surprise. But the King was not to prove a mere follower to Franco. After Franco's death, he astonished the nation by appointing as Prime Minister, Adolfo Suarez, who had spent his life working for Franco and seemed totally unsuited to dealing with a nation demanding change. But the King's choice proved correct; Suarez's knowledge of the dictatorship was to prove the very strength that enabled him to dismantle it. Change followed at a giddy pace.

Suarez was appointed in May 1976; by November he had persuaded the Cortes, stuffed as it was with old supporters of Franco, to agree to the establishment of a two-chamber parliament and universal suffrage. In December 1977 the change was approved by a referendum; the following spring political parties were legalized, including even the communist party. In June the first general election was held. Not surprisingly, Suarez's own moderate party was the winner — but not by an overall majority. To the nation's astonishment (and the disappointment of those who thought Franco's death would herald a revolution) Suarez then proceeded to negotiate a comprehensive series of agreements, the Moncloa Pacts, with all the opposition parties and then, when this was done, to draw up by consensus a new constitution defining Spain as a parliamentary monarchy.

While it was that act in October 1978



The Congress of Deputies takes cover as Lt Colonel Antonio Tejero Molina and the Civil Guard take over.

that marked the official beginning of the new Spain, it was two and a half years later that democracy had its baptism of fire. On 23 February 1981 (a date which all Spaniards seem to know), with Suarez still in power but his own party increasingly divided, a group of senior army officers decided that democracy had gone too far. The Congress and virtually all of Spain's top politicians were seized by a group of Civil Guards. In Valencia tanks took to the streets. But the attempted military coup was to fail, largely through the prestige of the King. By remaining firmly opposed to the coup and persuading military leaders not to back it, King Juan Carlos was able to maintain control and peace was restored in twenty-four hours.

The following year the party that Suarez had led was in such disarray that he himself left to form a new one. That left the field open for the Socialist Party to win the general election with the 40-year-old Feli-

pe Gonzalez as its leader. Just seven years after Franco's death Spain had an elected socialist government. Since then, Gonzalez has been skillfully piloting his party towards the political centre. After further success in the 1986 election he appointed the youngest cabinet in Europe, many in their early 40s and with a Minister of Industry and Energy who is just 35.

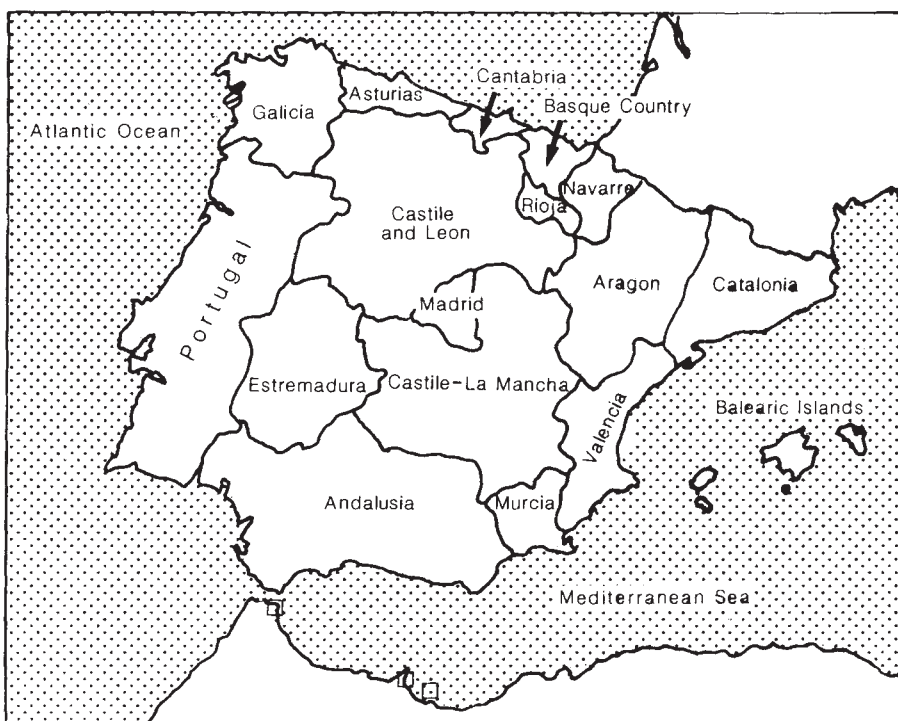
Alongside the establishment of democracy, other changes have taken place. Less well noticed abroad, but of great significance in Spain, has been the devolution of power to regional governments; Spain is half-way to becoming a federal state.

Devolution came in bits and pieces and was intended at first to provide for the demands of the Basque region — where violent demonstrations for independence had begun earlier — and Catalonia and Galicia; all regions with their own languages and culture. Spain now consists of 17 'autonomous regions'; each has its own ministers, president and administrative staff, but there similarities end. The Basques have gone furthest down the road

to autonomy. They collect their own taxes and send on to Madrid their share (a little over 6 per cent) of the cost of 'central' services — defence, foreign affairs, aviation and the like — while keeping back taxes for local government, including health, housing, social welfare and agriculture. Catalonia, in contrast, is content to let the central government collect taxes and receive their share. The Basques also set up their own police force and both the Basque Country and Catalonia have their own television channels and bilingual road signs. These changes have helped to quieten separatists, but the

problem remains; the main political wing of the Basque separatist movement has renounced violence, but its leftist military wing, whose own representatives are still able to win around 15 per cent of the Basque vote, has not.

Science, technology and education remain a disputed area in which clear and universal transfer of authority has not been arrived at. The Basques and Catalans are responsible for the running of their universities and schools, but in about half the regions the universities remain firmly in the hands of the central government, for the time being at least. No region has been able to grasp full financial control of its science and technology budget, although both the Basques and Catalans have their own regional science administrations. The unwillingness of the central government to transfer these powers is still very much an issue of argument. □



Spain is now divided into 17 'autonomous communities'; 15 on the mainland plus the Balearic Islands in the Mediterranean and the Canary Islands in the Atlantic. With somewhat different status are the two tiny North African enclaves of Ceuta and Melilla located, aptly enough, just across the sea from Gibraltar. Morocco would like the two towns back.

Science bodies await a reshuffle

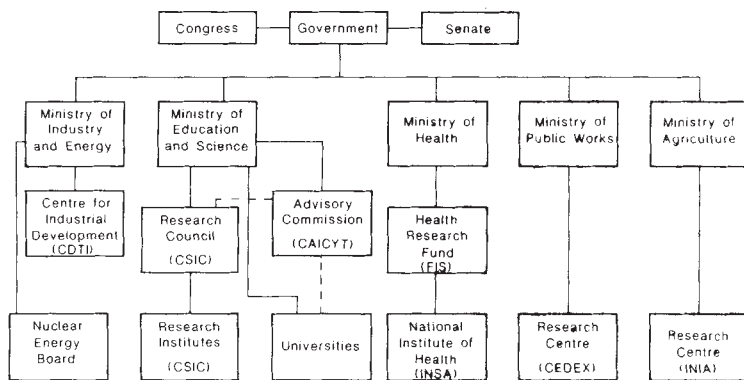
SPAIN'S science is not run quite like this any more (see chart) for the new law for the "Promotion and general coordination of scientific research and technological development" is just beginning to make itself felt. Under the old system the Ministry of Education and Science looks after both the universities, which have not traditionally been active in research, and the Consejo Superior de Investigaciones Científicas (CSIC), the science research council in whose various institutes research facilities are mainly concentrated. Grants for the research performed in both the universities and in the CSIC institutes are assessed by the Comisión Asesora de Investigación Científica y Técnica (CAICYT).

The Ministry of Industry and Energy is responsible for nuclear power development and for industrial development through the Centro para el Desarrollo Tecnológico y Industrial (CDTI), an organization advancing money for research where the risk is too high to attract ordinary investors. Medical research is funded through a separate granting agency, Fondo de Investigaciones Sanitarias (FIS). The Ministry of Public Works retains a large research centre, the Centro de Estudios y Experimentación (CEDEX).

Change came to the Ministry of Agriculture a few years ago when it lost direct control over

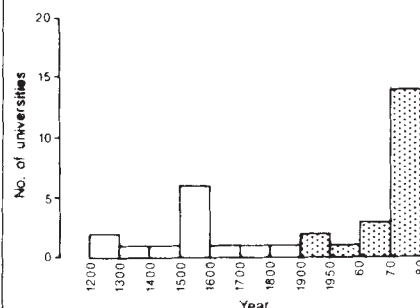
most of its research centres in favour of the autonomous communities in which they were located. But it retains responsibility for overall coordination and for Madrid's Instituto Nacional de Investigaciones Agrarias (INIA).

Under the new system the institutions survive but an attempt is made to improve coordination and planning through central control by the Interministerial Commission for Science



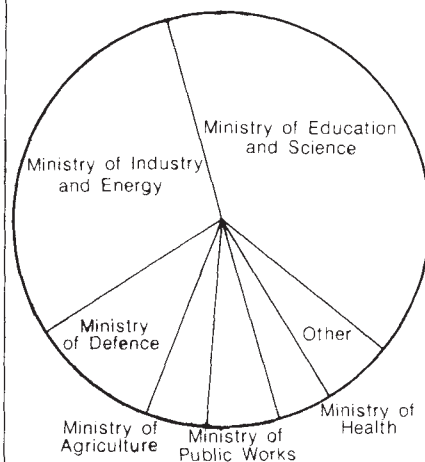
and Technological Development which contains representatives from all the ministries. This will sit at the centre like an octopus with its tentacles reaching out to all the research agencies. Its responsibility will be to programme all their activities through the construction of national plans. Advice on the design of the plan, its objectives and its progress will fall on an Advisory Council (a transformation of CAICYT). Quite how the new system will function is not yet clear for the Interministerial Commission has only just met for the first time, a situation perhaps symbolized by new charts showing everything joined to everything else by dotted lines.

Towards a second golden age



Why a second golden age? A glance at a chart showing when the universities were established shows that there has not been a boom like the past decade's since the sixteenth century. That was the time of the height of Spain's prosperity when riches were pouring in from the newly-colonized Americas (the Aztecs were conquered in 1519 and the Incas in 1532) and the King of Spain was also Holy Roman Emperor. Holland, Austria and much of Italy were under his control and Spain was the dominant power in Europe. Of the universities established then, Valencia, Santiago, Sevilla, Madrid (Complutense), Granada, and Zaragoza still survive.

Education ministry is the big spender



THE Ministry of Education and Science is the big spender among the government agencies with responsibility for research, even bigger than the Ministry of Industry and Energy. Total funds expended by the public sector reached 119,000 million pesetas this year — 80 per cent of the total. The small contribution of the private sector contrasts strongly with industrial nations like West Germany, the United Kingdom and the United States, where more than half of all research is funded by industry. The discrepancy helps to explain why the percentage of Spain's gross national product spent on research is so low, at most 0.6 per cent, in comparison with the major industrial nations which score from 2.2 to 2.8 per cent.

Law of science

More coordination from the top

"SCIENTIFIC research and technological development have traditionally taken place in Spain in a climate of apathy and lack of social stimuli, in the absence of the necessary apparatus to guarantee the efficient intervention of the authorities in the planning and coordination of the scant means available, with a lack of connection between the aims of research and the policies of the relevant sectors and, usually, between the research centres and producing sectors. It is hardly surprising, therefore, that Spanish contribution to scientific and technological progress has in general been small and inappropriate to the position Spain has occupied in other areas; and that, when this has not been the case, as in certain periods of the present century, the most worthy contributions have been the result of the isolated effort of particular individuals."

It would be impossible not to applaud a government that is able to set down such a

brutally frank appraisal of its own nation's shortcomings as that given above, taken from the introduction to the "Law for the Promotion and General Coordination of Scientific Research and Technological Development", passed by the Parliament earlier this year. Of course, the government finds it easier to make sweeping criticisms with the knowledge that the present defects can be blamed on the past dictatorial system. But even so the new law is a remarkable attempt to remedy the defects listed above, with the aim of ensuring that Spain can look forward to the day when it can speak on equal scientific terms with neighbouring countries.

The new law is not the beginning of the reform of scientific organizations in Spain. The ten years that have passed since the end of the Franco era "feel more like a century", as one researcher put it. The two largest concentrations of scientific ta-

lent, the institutes of the research council (CSIC), which have traditionally been the stronghold of research, and the universities, which have concentrated more on teaching, have both undergone considerable reform, as the following pages will make clear. What the new law really does is to reform the highest level of decision making and to bring it under the control of a new policy body answerable to parliament. That body is the Interministerial Commission for Science and Technology and it will be responsible for designing and overseeing the execution of the National Plan for Science and Technology.

The Interministerial Commission will contain representatives from all the ministries concerned with science and be chaired by the Minister of Education. It will have the power to decide the budgets, staff, and capital needed to carry out the national plan, to allocate funds to the various organizations involved in carrying it out and to coordinate the research activities of the different scientific bodies. Independent action by separate research organizations, each of them presenting its own budgets to the Ministry of Finance, will thus end. The plans of the commission will be debated by parliament and a yearly report on their progress made public.

It is not, however, the intention of the commission to impose rigorous control from the top over all aspects of scientific research. Within the national plan, there will be "National Programmes of Scientific Research and Technological Development" (in which information technology, biotechnology, and materials science are likely to feature strongly) designed by the commission. But the national plan will also incorporate "sectorial programmes" emanating from government departments and research bodies, and programmes from the autonomous communities. Overall control of the national supply of research personnel will be taken as the responsibility of the commission.

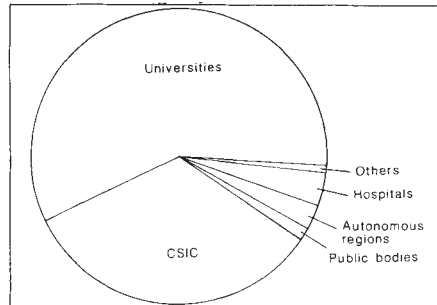
The status of a number of autonomous research bodies, including CSIC and the Nuclear Energy Board, will also be altered by the new law, principally so that they may more flexibly interchange staff.

Those who designed the new law feel proud of it; indeed they think the new structure could teach a thing or two to other countries, like the United Kingdom, which they see as very poor models for overall research planning. But it is much too early to say whether the Spanish system will succeed for, at the time of going to press, the commission will only have just had its first meeting. How quickly it will impose itself on the organizations it now commands remains to be seen.

Strangely, many in CSIC and the universities seem quite uninterested in the new law and doubt that it will bring about shifts in the direction of ongoing research. Perhaps more pertinent is the criticism

Sharing out the CAICYT cake

FOR the universities, CAICYT is the major source of research grants, although private foundations, industry and the Fondo de Investigaciones Sanitarias (FIS) of the Ministry of Health and Consumer Protec-



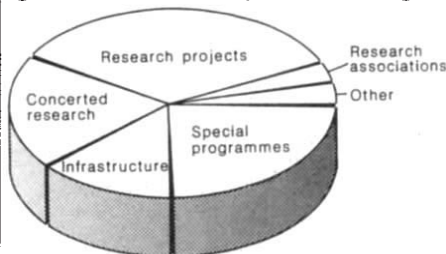
Distribution of CAICYT's project funds. A total of 3,826 million pesetas was given out in 1985.

tion are important. The latter provides much support to medical faculties and hospitals. Those working in research council (CSIC) institutes also have grant applications assessed by CAICYT but their grants can be topped up from CSIC's own funds.

Supporting research projects uses up the largest slice of CAICYT's budget, but still accounts for less than half of total expenditure. Special programmes take a quarter of the funds. They come in two varieties, mobilizing programmes which aim at the general promotion of a whole field of science (current themes are high-energy physics and biotechnology), and national priority research and development programmes which also promote coordinated action across a whole field, but aim at a more concrete economic objective. Both involve all research organizations, particularly the institutes of CSIC. Four programmes have

been launched in the latter category: biomass energy (which proved such a disaster that it was abandoned), aquaculture (which has stimulated the establishment of a new CSIC National Centre for Microelectronics with laboratories at the Autonomous University of Barcelona and the Polytechnic University of Madrid) and rail transportation.

Research in industry is promoted through concerted research programmes in which loans (up to 50 per cent of total cost) are given to private firms to cover their research costs. Part of the research may be totally funded if it is carried out at a government laboratory. If there are good



CAICYT's total budget — 9,100 million pesetas in 1985, but 13,000 million next year — is widely spread.

prospects for development of a marketable product then help from CDTI (see p.324) is likely to come in the later stages.

An increasingly large slice of the budget goes to improve research infrastructure, working through the autonomous communities. Smaller sums go to aid industrial research associations, to encourage foreign scientists to take their sabbaticals in Spain and to support international cooperation, scientific publications and congresses. □

Export market for advice

CAICYT is by and large a traditional research council: it assesses grant applications through the reports of external referees chosen by committees representing each subject area, it selects priority areas and arranges special grants to cover them and to aid cooperation with industry and the funding of research that might find application. All these activities and the other things it does (see opposite) are normal enough. What has made CAICYT a power in the land is that by all accounts it does its job very well and, more importantly, it was the first organization to bring an objective element into the assessment and planning of science. CAICYT's 300 or so referees have found themselves in demand. Grant proposals from both university and CSIC researchers normally pass before it (although if CAICYT chooses not to fund projects from CSIC researchers, they still have access to CSIC funds). And recently numerous other organizations have come to them for help, including some from abroad, like the Chilean research council.

The idea of making thorough analyses of the scientific situation really took off only recently. At the end of the Franco era one problem faced by both universities and the research council was that nobody really knew who was doing what or where and what equipment different laboratories had. Now CAICYT has a large computer and an enormous data base. A key word is sufficient to call up any scientist along with grant and publication record. Any piece of equipment can just as easily be located.

The data base has helped in putting forward realistic plans that Spain has the trained staff to meet, but CAICYT's concern for information sometimes appears to verge on the obsessional. The UNESCO coding system (now, alas, out of date) is used and there is a joke in the laboratories that a researcher who cannot find a 6-digit category to fit his or her work will have a short career. □

that without a real commitment to spend more money (which the government has not made) no profound change can be expected. What is certainly clear is that the success of the new organization will depend largely on the quality of the advice it receives. Responsibility for that will rest heavily with a permanent expert commission set up to advise the Interministerial Commission. Fortunately, considerable experience is at hand in the form of a body that has already proven itself a driving force in the scientific scene — the Advisory Committee for Scientific and Technical Research (CAICYT) of the Ministry of Education. The new permanent commission will largely be the old CAICYT with a few additions and new name. □

The research council

Back on the road to recovery

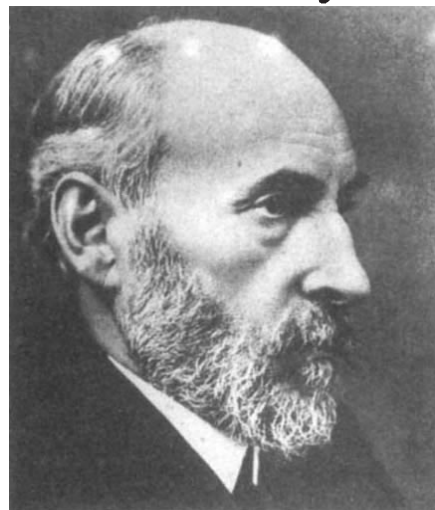
THE Consejo Superior de Investigaciones Científicas (CSIC), the nation's largest multidisciplinary research body, and its traditional source of scientific strength, is coming through a profound reorganization. At the end of the Franco era the CSIC had fallen so low that there was talk of abolishing it altogether. But instead it was reformed, more vigorously than the universities, and is reasserting itself both in basic and applied research.

The CSIC of the 1970s could provide a lesson in how not to run a research organization. The number of its institutes had grown to 170. Many of them were extremely small and had been set up for personal rather than scientific reasons; some contained just one person, for the old style Franco-era professor was a strong figure who found it hard to work under anyone else. Salaries were low and laboratories poorly equipped. The staff was growing older; by 1981 the average age was 48. But after the restoration of democracy things began to change and by the 1980s the pace had increased rapidly. Now just 76 institutes are left, some that vanished were closed down, but most were fused with others. That meant many former directors had, once more, to become subordinates or to leave.

Some were dispatched by lowering the age of compulsory retirement from 70 to 65. Not all went peacefully. Here and there in reorganized institutes there are autocrats who have hung onto large areas of laboratory space, despite the entreaties of their younger and more active colleagues. Elsewhere, quarrels over who was to gain power have left people no longer on speaking terms. And despite reorganization, integration of laboratories is sometimes proceeding in name only, but that perhaps requires only the passage of time to cure.

A flood of new appointments allowed young people to enter CSIC, or to move up its hierarchy. In 1981 there were 1,100 scientific staff; by the end of this year there will be 1,800. Over the same period the budget doubled and CSIC made efforts to open itself up to the universities and to industry. Fourteen joint centres have been set up in the universities, usually the university provides the building but all the laboratory equipment and around a third of the staff are provided by CSIC. Some 40 per cent of the Consejo's research projects also involve university researchers, for although they do not have direct access to CSIC's funds, they can make joint applications with CSIC researchers.

Industrial contracts are being signed at the rate of three a week. This year a little over a fifth of CSIC's total budget will



RAMON y Cajal, who won the Nobel prize in 1906 along with Camille Golgi for establishing the neuron as the unit of the nervous system, signed, as its first president, the document that brought the Junta para Ampliacion de Estudios e Investigaciones Científicas into existence in 1907. The Junta was to prove a liberal scientific institute sending many people abroad for training and leading a short scientific renaissance. Then the civil war came and in 1939 Franco suppressed the Junta. It was soon rebuilt as the CSIC, but not with the same people — some three-quarters of Spain's senior scientists had fled into exile. The new institution was given the quasi-religious goal of "rebuilding the unity of science broken by the philosophy of rationalism", but retained some liberal traditions; in the Franco regime's view, intellectuals were safer in the CSIC, where there were no students for them to teach, than in the universities.

come from industry. Industrial researchers can also come to CSIC laboratories for training — although it is not yet possible simply to send CSIC researchers to industry. CSIC's head, Enrique Trillas, is particularly proud of the successes of the Office for Evaluation and Technology Transfer and that industry is now coming to the CSIC without being asked.

Much then has been achieved. Even the Napoleonic style examination — the *oposición* — has been altered, as it has in the universities (see p.320). Now, to get a first post within the research council, there is a public examination of the candidate's academic record and research speciality. But later promotion depends only on an assessment of the researcher's work. More important is that in both cases it is no longer a jury drawn from CSIC institutes only that makes the decision, but one containing professors from the universities and, increasingly, from abroad. In the past juries often came from the candidate's own institute, making the promotion of internal candidates and the exchange of favours a common feature of institute life. The rejuvenated CSIC institutes still span

the whole of science: from history, through molecular biology to construction and cement.

For many in these laboratories, the greatest change has been that research proposals are now thoroughly assessed. In the old days a grant came with a permanent position. Now researchers receive a salary and basic infrastructure and must apply for grants, either from CAICYT and CSIC, private agencies or industry. And all scientific evaluations are made by CAICYT (see p.317) and its teams of referees, rather than CSIC itself, although an economic evaluation is first performed



CSIC's head Trillas.
by a committee from CSIC.

The great changes made in CSIC do not mean, however, that its troubles are over. Although CSIC has better facilities for research than the universities and is the place where most research is done, it is still tiny, with less than 2,000 researchers, for a country the size of Spain. And more important, scientific traditions are not made overnight. The best institutes are at international levels. But they are few and far between compared to the run of the mill institutes that rarely produce a paper for an international journal. Moribund institutes still need infusions of talent, particularly of people trained abroad, but returnees face no easy task at institutes where excellence is not highly valued.

Most CSIC researchers, although happy to moan about the problems that remain, recognize the enormous improvements of the past ten years. If anything, what really needs to change is the attitudes of researchers themselves. One quote (delivered after a half hour of complaints about the bureaucracy) sums it up: "Our facilities are really no worse now than in the United Kingdom, and we have no lack of funds if we apply for them. It's up to us!"

Fundamental changes in attitude will come as talented young people take power within the institutes, coupled with increasing competition to gain grants. Then all that will be needed is a several-fold expansion of the system, before the research council begins to look like a ranking European scientific organization. □

Science evaluated

Still some catching up to do

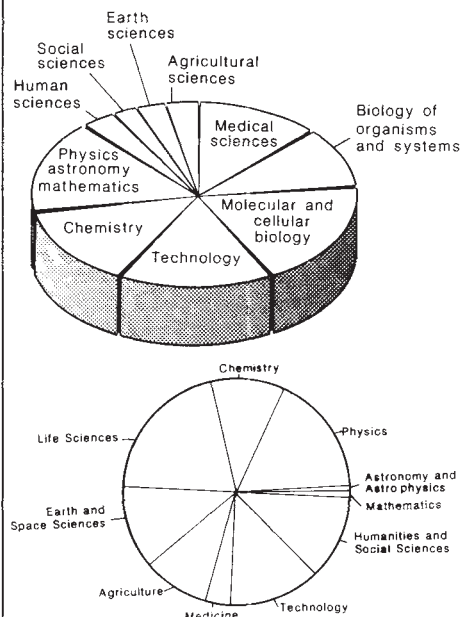
It would be a mighty labour to try to give a complete and fair evaluation of the strengths and weakness of the different fields of science in Spain. Even in fields which are generally poorly developed there may still be outstanding characters making important international contributions: in ecology, for example, in which Spain produces few publications, there is Ramon Margalef at the University of Barcelona whose books have become standard works in quantitative ecology. Sweeping generalizations would not do such individuals justice. Something of Spain's record can, however, be read from the distribution of research grants of CAICYT and of research projects by CSIC coupled with the store of gossip available at every Spanish laboratory.

There is general agreement that molecular biology is strong (witness CBM, p. 319) and is attracting sizeable funding; "biochemists have had a powerful lobby for years" as one researcher put it. Physics has fared very much worse, with no large experimental facilities, and having seen people of the calibre of Blas Cabrera vanish into exile after the civil war. Only theoreticians managed to maintain standards. That should now change as a result of Spain rejoining CERN and an accompanying 'mobilization programme', in high-energy physics.

In nuclear physics there are some good Spanish scientists abroad but poor facilities within the country. European collaborations offer hope. Solid-state physics has also suffered from lack of equipment but this has already started to change.

Laser physics is the subject of a special CSIC mobilizing programme but a start is

having to be made from almost nothing and standards are still low. In mathematics young returnees from foreign universities are just beginning to make things move and some specialized areas are already at international level. Astronomy



The share-out of research funds from CAICYT (top) and CSIC's project distribution.

appears to be a major disappointment. Despite access to the superb international facilities in the Canary Islands, very few Spanish papers appear in top journals. Atmospheric physics, earth sciences and whole organism biology remain very old-fashioned with only a few exceptions. But in geology, efforts are beginning to be made to link up with foreign researchers.

Through English rags to scientific riches

SOMEWHAT over half a century ago, Roman Areces was working as a messenger boy in a little shop in Havana, Cuba. Now, through the foundation named after him, he is the biggest private source of support for scientific research in Spain. The path to philanthropy came through 'El Corte Ingles', a store he set up in Spain to make and sell suits with 'the English cut'. The name has stuck, although his department stores now sell everything from asparagus to zippers and it is impossible to find a city centre without one. And he is still selling suits — his factories producing ready-mades are the biggest in Europe. Overall, the group is one of Spain's biggest and parts of it have even gone into high technology — including electronics and biotechnology (see p. 325).

The foundation was set up ten years ago when Roman Areces, aged 72 and still a

bachelor, decided to use part of his fortune to aid Spanish science. By international standards the foundation is not enormous but it is not so small either: In 1984–85, 147 million pesetas — close to a million dollars — were given out in grants along with double that amount for conferences, publications and to help equip research centres. Each year four research areas are chosen for support, and these replace (or are added to) the previously chosen areas. Anyone working in appropriate areas can apply for funding and grants are awarded on the decision of panels of referees. Research topics chosen recently range from robotics and solar energy to ocean sciences and cardiovascular disease. Individual grants are quite large — 23 million pesetas for a robot vision system, 13 million pesetas for molecular study of calcium channels in chromafin cells are examples. □

Centre for Molecular Biology

Long queues for laboratory space

"The trouble with the Centro Biología Molecular," said one research council administrator, "is that everyone wants to work there . . . but it's impossible". The 'CBM' has gained an international reputation for excellence and a national reputation for being far and away the best molecular biology research institute in Spain. Plus it has in residence Spain's Nobel prize winner Severo Ochoa. That has made it a popular place to be — a condition not without difficulties. Many at CBM are frustrated that they cannot accommodate all the researchers they would like while the research council would be happy to see some good scientists (particularly those returning from abroad) building good groups elsewhere rather than adding to CBM's lustre.

Among the main lines of research are virology; DNA regulation and protein synthesis; and molecular neurobiology, including the biochemistry of ageing and Alzheimer's disease. Perhaps best known (at least to readers of *Nature*) is the strong tradition in the developmental and molecular genetics of *Drosophila* begun by Antonio Garcia-Bellido on his return from Caltech and continued by Ginés Morata, Pedro Ripoll and Juan Modolell.

What makes the CBM interesting to the outsider is not just the science (of which there is plenty) but the way in which it has overcome some of the weaknesses in Spain's research system. It is the only major institute to have ended the traditional separation between CSIC institutes (specializing in research) and the universities (specializing in teaching). Under one roof (the university's) are the twin molecular biology institutes of the Autonomous University of Madrid and the CSIC.

The two sides share administration costs and elect their own representatives and a joint director. Those from the university side still have teaching duties while their CSIC colleagues have none, but that is the

only difference and "we live in peace" as the director Jesus Avila put it. Both sides provide somewhat over 30 staff, and the CSIC over 60 'technical assistants'. Seventy-five research students complete the numbers. Fusion took place in 1975 before the need to bridge the gap between the universities and the research institutes was so widely recognized. But despite its success, the CBM has no imitators (only one institute has a similar structure). That might be because it took friends with political clout to set up the institute (among them Federico Mayor, then in high office). Their presence helps explain why there is no shortage of good equipment.

Organizational changes have helped overcome one common problem of the research institutes — a poor laboratory infrastructure (stores, animal rooms and the like). CBM researchers agree to apply jointly for a core grant as well as obtaining their own individual grants (although some researchers might end up with less this way) to maintain the research facilities. Elsewhere there is usually a levy on individual grants to provide common stocks but it rarely proves sufficient.

The core grant also enables them to tackle jointly one bizarre curse of CSIC laboratories: the technical staff belong to a large public service union and work from eight in the morning until three in the afternoon, so just when things are in full swing it is time to get the bus back to Madrid. Some extra money can be provided to keep some of them on later. And other problems, like keeping the heating on when the university switches it off (when classes are over) can be dealt with.

But despite these successes a gigantic problem remains to ire CBM researchers — the lack of space. There just is not room for everyone. "Asisnine" is how Dr Avila described the lack of coordination between the research council and the facility that has led to "at least eight people being employed who can't work here because they do not have space". A part of the problem is that space is not always distributed rationally. And that might be because, as a visiting advisory committee of foreign experts pointed out, the director has no real power. He or she is elected for two years which, according to Margarita Sallas, often seem "lost years from a scientific point of view, just a lot of administrative work". It is never easy to find some-

one to take on the job.

Without real power at the top, space cannot readily be allocated or projects terminated. Newcomers speak of having to bargain with territory holders to find a corner. As Juan Modolell put it "[space] is not dependent upon scientific accomplishment. I really have this laboratory because of historical circumstances, other people have half my space, or no space, and others several times more, but it has nothing to do with ability". That in turn, he



Space is at a premium here in Margarita Sallas's laboratory as elsewhere in the CBM. The record seems to be ten people (and a few million fruit flies) in a room of just 49 square metres.

says, reflects a more general problem, that "there is very little policy of promoting excellence".

The director, Jesus Avila, does have plans to try to readjust laboratory space, if he is able (he was last seen by this correspondent wearily threatening to resign if promised administrative help did not materialize soon). What he wants to see is an institute with a potential for growth, that can let in young people. "Every institute has to grow or die, and we want this place to grow rationally", he says.

A more complete solution than readjustment can achieve might come with the new National Centre for Biotechnology to be constructed close by. There are doubts about just when it will be built (the decision was announced in 1984 but construction has not begun), but excellent molecular biologists are what it will need. There are many at the CBM who could do the job. A danger though is that the new institute will turn out to be just an annex to the CBM (or a place for some of its senior staff to retire to) instead of what it is meant to be — a laboratory working closely with industry. Those in industry who want something different from the CBM's strength in basic research complain that few people with industrial experience have been suggested for the board. For the researchers perhaps, as Juan Rojo, Secretary of State for the Universities and Research put it (very gently) "there may have to be a little shift of centre of gravity towards problems of interest to industry... a *Drosophila* group, if it's big, maybe some of them would shift a little to vaccines, agricultural products and so on." □



Severo Ochoa has now made his home at the CBM. He left Spain in 1936 and eventually became an American citizen working at the New York University College of Medicine. In 1959 he shared the Nobel prize with Arthur Kornberg for isolating enzymes that made it possible to synthesize RNA and DNA.

The universities

Worst problems overcome

Ask a group of professors about the state of the universities and you will hear a bewildering mixture of opinions. But after a little while a pattern emerges; praises come for the enormous improvements that have been wrought in the university system in the space of a few years but savage criticism is reserved for the defects that remain, often coupled with pessimism that they can ever be fully cured. Whether more criticism than praise is heard depends partly on the personality of the speakers, and very much on the amount of time spent abroad.

While perfection may be an impossible goal it would be too much to agree with the several cynics who said, "We were right at the bottom, so we could only get better". Planned reforms have had their effect and conditions for researchers are improving rapidly. The question is whether the reforms really went deep enough to give the momentum for the continued changes needed to bring Spanish universities up to the best international standard.

After the massive growth of the universities towards the end of Franco's rule real reform came with a new law for the universities — the *Ley de Reforma Universitaria*. Its overall aim was to remove the autocratic control exercised by central government over the universities, which made all universities equal and removed any spur to excellence.

Under the new law the universities were given power to write their own statutes, with the approval of the government, through the action of a University General Assembly (*Claustro*) freely elected by professors, students and administrative staff. Departments were rationalized. A huge number had been created under the old regime; many, it seems, to satisfy the personal ambitions of an individual professor or to separate those who had quarrelled. Resources are now concentrated in single departments for each subject area.

The enormous problem of untenured teachers was also tackled. The rapid rise in student numbers during the 1970s had made it impossible for the universities to appoint sufficient staff through the usual bureaucratic channels and many lecturers were appointed on short-term contracts. By the end of the boom there were more of them than there were tenured professors. Commissions were set up under the new law to decide who among the non-tenured staff should be taken on as professors; not surprisingly, perhaps, difficult choices were often avoided and the great majority were found suitable.

With their own charters written and

Research conditions still lagging behind

THE University of Barcelona's School of Biological Science boasts an impressive piece of modern architecture, full of complex curves on the outside and with a huge central well inside, around which stairs zig-zag upwards. But the building was planned in a more autocratic era and little consideration was given to the needs of those who were going to use it — with the peculiar result that corridors pass through the middle of research laboratories and libraries while unusable empty spaces abound.

Space is one thing the school is short of, for the study of biology attracts vast numbers of students; 3,000 in all. With numbers like that, some of the greatest difficulties faced by the university teaching system can be clearly seen at the school. There is no selection of students entering the school (biology is often a second option for those who wanted to study medicine, one of the few courses where numbers can be limited) and after one year only 800 of the 1,500 who entered are left — a drop-out ratio also seen in many other places.

At the end of the five-year course, half of the graduates will be unable to find a job, and only 10 per cent of them a job related to biology. With these prospects, it is not surprising that just half the students complete the course in five years; others retake courses they failed *en route* a few times — some stay on ten years and are well into their 30s. Only 10–20 per cent of the stu-

dents have grants so the prospect of money running out does not provide a spur.

At the start, classes are big (more than a hundred students in each) and there is no chance of tutorials or seminars, nor much practical teaching — making the courses almost 'theoretical biology'. Only in the later years can groups of 20–30 students be dealt with. Given these factors, few students are highly motivated, and the long



Impressive architecture but does it work?

teaching hours make a real research effort from the staff difficult to attain. On top of that, back-up services are almost nonexistent.

Professor Jaume Baguna, who heads the Department of Genetics, has a dozen researchers and the same number of doctoral students without a single technician and only a part-time administrator. As Professor Baguña says, "things have improved and change is coming, very slowly. But university research conditions are still very bad compared to the rest of Europe". □

Industrial contracts provide rich harvest

No less than 20–25 per cent of the entire budget of the Polytechnic University of Catalonia, some thousand million pesetas, is earned by research projects with industry, municipal government and hospitals. The added autonomy given by the reforms in university law was a great help according to the rector, Professor Gabriel Ferraté, not because it granted true autonomy to the polytechnic but because it gave the management the flexibility necessary to find external sources of funding. The polytechnic's income from research contracts

is now increasing almost exponentially.

The polytechnic was one of those in the camp opposed to keeping the status of teachers as civil servants, according to Professor Ferraté. But under the new law, winning contracts can also mean an increase in salary and that helps to provide motivation, which the polytechnic is taking full advantage of.

Research at the polytechnic has been buoyed up by contracts with young and energetic Catalan firms in the new technologies. Although the school of civil engineering has a high reputation, real glamour, as elsewhere, has passed to computer sciences, computer-assisted design and manufacturing and telecommunications: successful projects have been run on fast pattern recognition systems and automatic map-making from aerial photographs. Professor Ferraté himself founded a hybrid Institute of Cybernetics combining polytechnic and CSIC facilities. Its main thrust is "artificial intelligence applied to robotics", and several cooperative projects are running with US universities. □



Ferraté: looking to intelligent robots.



Comfortable surroundings and a strong Jesuit teaching tradition make Comillas popular.

Comillas can set high entry standards

ONE advantage of being a private university is that you can select your students — and to enter the Universidad Pontificia Comillas, run by the Jesuits, is by no means easy. Exam grades have to be very high and are followed up by psychological tests and an interview. Part of the reason for the university's success is that its students almost always find jobs. "People trust our recommendation", says the university's rector Guillermo Rodríguez-Izquierdo. Many applicants are the children of graduates; it is apparently a painful process to turn away those of them that are not up to the mark academically.

The university, now in Madrid, began

as a seminary set-up at the end of the last century and its degrees are recognized by the state following an agreement between the Holy See and the Spanish government. Business administration and economics are now the most sought-after courses for the university's 6,000 students, but there is a popular course in engineering and a unique system by which industry supports doctoral students.

The university is aiming to achieve a different style from the others. "We try to combine science with radical concern for human problems", says the rector. And perhaps just as important, "students come to the classes on time!" □

Getting basic research on the rails

THE Autonomous University of Barcelona must be the only university in Spain to have its own purpose-built railway terminus. That is because the university was built, in 1968, 20 kilometres outside Barcelona with nowhere nearby for its 28,000 students to live. The university's newly elected rector, Professor Ramón Pasqual, readily admits that this has put the university at a disadvantage and that good students may choose to go elsewhere. But he is one of the optimists about the university system's future, and of the ability of his own university to offer something special. "Under the new law," he says, "the most important thing is that you can be different. In a few years there will be good and bad universities and the Ministry of Education will play differently with different universities. Some will go well, others will have problems".

New joint CSIC university institutes, including those for microelectronics and for materials are being set up on the campus and a science park is appearing across the road (Catalonia's 'Silicon Valley'). Links

with industry are growing.

Pasqual is also optimistic about Spain's basic research effort. "Now every journal has a paper from a Spanish researcher", he says. And as a theoretical physicist he applauds the decision to join CERN and the European Synchrotron Radiation Facility which will provide completely new opportunities for experimentalists. The big problems that remain? "Universities are still very poor. We lack technicians and we need administrative support — departments here still have no secretaries". □



A 'campus university' in the true sense.

approved what will happen to the universities? Will each university acquire its own distinct character rather than being one unit of a centralized bureaucracy? Will the best researchers (or teachers) find themselves in demand from competing universities? A glance at the law suggests that the answer will not be simple. It is a remarkably large document — some 59 articles and 22 pages. The autonomy given to the universities is of a form which it prescribes very exactly. All are now autonomous — but in the same way.

A key problem, not fully resolved, is that it is hard to reward directly those who are talented. There are only two full-time positions; the full professor (*Catedráticos*) and assistant professors (*Profesores Titulares*). Salaries are fixed. There are no grades, and there are no easy means to increase a professor's salary (other than the fixed annual increment). This has meant there has been no means whatever to try to attract the brightest and best researchers, nor any reward for creating excellence. The result, over the years, has been an almighty indifference to the quality of research and teaching at a university; as one young researcher at the Barcelona Autònoma put it, "no-one really cares if their university is good or not. No-one ever fights to get the best man. They are happier with someone worse who won't provide any competition". And without any incentive to find good people the tendency has been for those already in place simply to rise through the ranks. The key step to a successful career has always been regarded as forming a close attachment to a powerful professor early on. The new law does allow researchers to double their salaries through contracts with industry, an incentive that many universities, particularly the polytechnics, are grasping with vigour. But the law provides little stimulus for basic research.

While autonomy may not have gone far enough in salaries, many say it has gone too far on appointments. The extra freedom given to the universities to decide on their staff has not always been wisely used — although it is still far from complete freedom.

All university staff are civil servants and in the past appointments were centrally controlled. Each required one of Spain's great official rituals — the *oposición*. In a series of public examinations candidates had to prove their worth, sometimes in a direct debate with their competitors, before a tribunal of seven appointed by the government. The range of questions that might be asked was huge; it was not uncommon for candidates to be given a topic and be required to deliver a lecture on it the next morning. Nowadays, things are much simpler and the procedure is closer to a formal panel interview than a public ordeal. The examination comes in two parts; the first on the candidate's curricu-

lum vitae, and the second on the field of research. The experience can still be harrowing, as much for the tribunal as for the candidate. When there are many applicants each has to pass through the same lengthy procedure and a tribunal can be busy for a month.

Where the university has gained control is in the selection of the tribunal. Two of the five members are provided by the university that has the vacancy; the remainder are chosen at random from among professors in the same field. At the same time the university may provide a clear 'profile' of the person they want to fill the job; the combined result is that it has become much easier for the universities to choose who they really want. And what it seems they have all too often wanted is simply to promote someone in the same department. Greater autonomy has, if anything, increased inbreeding rather than fluidity among researchers.

There are even stories that universities have connived to avoid a regulation that says no one should be appointed who has not spent at least some time at another university by making linked appointments in which professors apparently move but in fact stay where they are. And these problems come at a time when there has been a boom in new posts.

Just how bad an effect this has had is very much a matter of opinion; some go as far as to say that the net result of having to give many untenured teachers tenure, alongside misuse of the selection system, has filled the universities with mediocre young professors who will now dominate university life for thirty years. Others say these are just teething problems. And some politicians blame the rectors for the universities' problems. Rectors are elected under the new law and this makes them anxious to please their constituents rather than make strong decisions. Cer-

Complutense tackles the problems of scale

THE Complutense in Madrid, with 118,000 students and 5,000 professors in 18 faculties, is one of the biggest universities in the world and by far the largest in Spain. One in five of all students are educated here. Managing a university on this scale is "often terrible; each small problem is multiplied by hundreds", says the vice-rector, Professor Angel Martin Municio.

In research, new measures are being taken to break down the barriers between disciplines and between the many small empires which have traditionally formed in the universities. The carrot being offered to persuade groups to get together is expensive technology: a world class nuclear magnetic resonance imaging centre is being set up to serve all researchers in the university, for example. Another success has been in bringing together all the relevant groups in the six hospitals attached to the medical school by providing funds for animals for use as disease models. This way the university also gains facilities good enough to allow participation in international programmes.

Headway also seems to be being made

with the problem of student wastage; some departments in some universities lose half their students after one year. This year, within Madrid's three universities, students applied for a subject rather than a university. The result is that 80-90 per cent are studying what they want and motivation has been improved.

As elsewhere, the Complutense is actively pursuing contacts with industry. In five or six years, Professor Martin says, "Uni-



Professor Martin needs a helping hand.

versities will be different according to their courage and effort". □

tainly, many in the political world now regret that rectors are not appointed.

Reform could have been carried much further by doing away with the automatic appointment of staff as civil servants with jobs for life. There were many, including very senior staff, who fought to replace lifetime tenure with long-term contracts to enable selection to be repeated at intervals. Curiously, the socialist government was of the same mind, before it faced the challenge of putting its ideas into practice.

One way of quickly raising standards of research is to send people off to the

world's very best institutes for several years of doctoral or postdoctoral training. Spain has certainly not ignored this option and several exchange agreements and fellowship schemes exist. But those who go abroad say their problems begin when they try to come back. Foreign degrees are not automatically recognized, nor can a returning scientist expect an automatic right to the same amount of laboratory space and access to equipment as those who have built their territories at home. The administration is aware of the problem — and some who have returned successfully say that those who have not do protest too much. All want to work in the very top laboratories, ignoring the importance of establishing groups in the smaller universities.

One part of the university reform that seems to have had little impact is the decision to transfer responsibility for the universities to the autonomous regions. In practice this has not meant universities acquiring a strong regional flavour. All that has happened is that administration has been made somewhat easier, for not all requests have to go as far as Madrid. But Madrid still holds the budget strings and transfer has not been total. Half of the universities (including, naturally, those based in Madrid) have simply chosen to leave things as they are. And the grass always seeming greener on the other side of the fence, those universities that have been transferred protest that Madrid universities are able to collect more money, while the Madrid universities claim the others collect funds both from the central and the regional government. □

Autonoma tops for molecular biology

THE campus of the Autonoma in Madrid is instantly recognized as 1960s; bits of decaying abstract sculpture dot a park surrounding concrete buildings connected with tunnels, perforated by underpasses and decorated with murals symbolizing something that has long passed out of common knowledge. Just as at an airport you know what kind of place you are in — even if you are not sure of the continent (except here, perhaps, for the rare fragrance from the surrounding olive groves). And like an airport, the biggest facility is the car park; the campus is a half-hour ride from central Madrid and there seems to be an endless line of students trying to hitchhike a ride back there.

But being close to Madrid has helped the university attract some of the country's best researchers. The Centre for Molecular Biology is far and away the top institute in its field and there are strong groups in



Molecular biology and sculpture predominate.

chemistry and in physics. The new Institute for Biotechnology is to find its home here. □

Science policy

The man in the hot seat

PROFESSOR Juan Rojo, a physicist, is Secretary of State for Universities and Research, which gives him overall responsibility for both the university system and CSIC. It is a job which he describes as "bad in the sense that I have to deal with personnel problems but good in that I can contribute to making stronger links between the universities and the research system". Professor Rojo's horizons will soon widen more than a little: once the new Interministerial Commission is running he will lead its advisory committee and that will give him a chance to look at the government's total 100 thousand million pesetas spent on science.

Professor Rojo seems optimistic about the universities' future. To those who believe that giving increased autonomy to the universities has actually increased their "inbred nature" he says. "Some, it is true, have used their autonomy to promote their own people but if you give autonomy you cannot control the universities at the same time. I believe autonomy will open the door for excellence." He is proud of another aspect of the law of university reform that is not widely commented on: the creation of the "Social Council", which contains 12 people from outside the university (including representatives of industry and the trade unions) and eight from inside. This "provides a counterbalance to autonomy", and he says "we have good evidence that it is fighting inbred-



Professor Juan Rojo wants stronger links between universities and the research system.

ing". The council has powers "to increase the salaries of professors of special excellence, and to decide whether vacant chairs should be transferred to new areas . . . or, say into two lectureships".

Professor Rojo is also tackling another source of complaint: that Spanish researchers who have worked successfully abroad find bureaucratic obstacles prevent them returning easily. First task is to make a standard list of centres from which degrees will be recognized, "so that people don't have to waste time translating their theses" (for doctoral degrees and so on to be locally validated). And "second,

Agricultural research

Molecular biology in the pig-pen

AGRICULTURE is the only sector in which most research centres have been transferred to the autonomous communities. A precedent for a fuller devolution of research was not set, it seems, because agriculture is recognizably different. Each area has its own agricultural problems and needs its own research institutes to tackle them. Their activities are, however, closely coordinated from the centre by the National Institute for Agricultural Research (INIA).

Under the new Law of Science programmes for agriculture remain the responsibility of the Minister of Agriculture, but they are complemented by more wide-ranging interdisciplinary programmes that include other ministries and are parts of the national plan. Programmes are monitored by a commission from INIA containing referees from independent organizations.

One research institute, in Madrid, is directly attached to INIA (and takes its name from it) and is the home of most of agriculture's high-level basic research, including work on one of Spain's biggest agricultural problems, African swine fever. Spain and Portugal are the only countries in Europe where the disease is widespread. That means that although pigs are Spain's biggest agricultural earner (twelve million head a year are consumed), exports are almost entirely prohibited for the disease spreads easily, particularly in contaminated pork, and is usually fatal.

It was the rise of air travel that brought the disease to Iberia. Pigs kept near Lisbon airport, fed on swill from Angola, were the first to be infected. From there the disease spread into Spain in the early 1960s, before much was known about it. Mass slaughter of infected animals has proved the only method of control. When the disease reached the Dominican Republic in the 1970s the drastic decision was taken to slaughter all the pigs in the country. Small outbreaks in North Europe have been contained by the same means.

In Spain slaughter of infected animals and strict control of movement of stock mean that eradication may be possible in a

few years. But now the infection has spread into the wild population, periodic re-infection is probable. The only answer, it seems, is an effective vaccine. But the search for such a vaccine has often seemed a hopeless task for the virus does not stimulate production of neutralizing antibody, a property shared by one rare mammalian virus; that causing Aleutian disease in mink.

Work at the institute now concentrates on understanding the basic molecular biology of the virus, immunological reactions to it, and its infectivity in pork products. Study of the virus genome does not encourage hopes that a vaccine can be produced easily. A variable region containing several gene families has been identified which suggests that the virus escapes the immune system by constantly changing. Now it is important to find out exactly what the genes code for.

The immunologists are a little more hopeful. Some pigs do recover from the disease and overall mortality has decreased over the years, suggesting that the immune system can do something. Some experiments suggest that the virus escapes from the host's immune system not because it is changing but because viral proteins are masked by host proteins. There are now hopes that there are some small virus-specific proteins that are not masked and against which antibodies can be produced. Serum from pigs that recovered from infection has already been found to inhibit infection by the virus *in vitro* tests. Can this encouraging result be built on in the future?

What the researchers now need ("worth their weight in gold", according to the institute's director Dr J. Sanchez-Vizcaino) is pigs that are serologically positive but healthy. That and another couple of years work will show if there is any hope of a vaccine. Research will soon be intensified with a move, 40km from Madrid, to a new Centre for Animal Infection, one of the largest in Europe. It will be perhaps the only one to have a large part of its floor space dedicated to pigpens along with P3 and P4 laboratories. □

we are trying to favour the kinetics of return. We are encouraging people to take their sabbatical leave here, as well as giving 2-year fellowships".

On a wider issue, work for the new Interministerial Commission has already begun. Themes under discussion for national programmes are information technology, biotechnology, new materials, engineering sciences and food technology.

Although many feel that Spain's total research expenditure is shockingly low at 0.6–0.7 per cent of gross national pro-

duct. Professor Rojo points out that the rate of growth has been high, coming up from 0.4 per cent in the past five years. The top priority now is "investing in the training of people for the future. We now have 4,500 people at the postgraduate level which is reasonable". But there is still a serious gap to be filled in the training of technicians. "For people of 18 there is very little alternative to going to university, there are practically no technical colleges at all and we need to develop the level of technical training." □

Industrial research

Companies need more science

SPAIN'S entry into the EEC, on 1 January 1986, was the ultimate proof that the introspective Spain of the Franco era was dead. But enthusiasm for the wider horizons of Europe has to be balanced with the economic realities — Spain's industry now faces its biggest challenge for decades.

The key problem is that industry has not invested in research and development and that has left Spain with a pitifully small total research expenditure — just 0.6 per cent of gross national product, against the 2.4 per cent of Britain and the 2.7 per cent of West Germany — and largely low-tech industries. Boosting the research figures is now one of the government's top priorities.

Every research body is busy seeking out links with industry. The research council (CSIC) can boast that it is signing three contracts a week with industry on behalf of its institutes: and the polytechnic uni-

versities are doing particularly well with income from industry rising exponentially in some. Researchers are clearly under pressure to push industry, but is industry really prepared to pull? Dr Joan Majó, until a few months ago Minister of Industry and Energy and an electrical engineer, gives his views.



Former industry minister Joan Majó sees Spain as a major food-producer in years to come.

“There is not enough research in industry, that is clear. People simply blame the industrialists for this but it is more the fault of the circumstances. For years we have been living in a closed market, without external competition and there has been no need for R&D to make money. Some companies with external trade recognize the need and have reached good technological levels. But we do not have any very big multinationals that can support huge research efforts. Instead we have many medium-sized companies.”

Can industry survive EEC entry and reach European levels? “We are joining at

the best moment in the last ten years. Why? Because the opportunities are now greater than the dangers. Five years ago it was different. Industry had still not reacted to the industrial crisis of the 1970s and there was overproduction in Europe for which Spain could have provided a new outlet. Change began in the early 1980s and now we are able to compete. Productivity and investment have risen although some sectors, like iron and steel, have special problems.”

Dr Majó believes that in any case Spain does not need to do all its own research. “The problems of the use and generation and technology are not the same. We have to accept that part of the technology we need is not generated here, or is developed by external firms. External firms should be encouraged where Spanish ones have no chance of success in the short term. The key is to be selective.”

A range of familiar incentives — tax breaks, import restrictions and direct grants for research — are being used to

stimulate industrial research. Some 35,000 million pesetas went in direct grants to electronics companies during Dr Majó's own period in office from 1982 to 1986. But there is still a lack of research infrastructure and trained personnel. One danger is “a possible collapse of the supply of technicians caused by the small scale of the technical universities. If major industries did all set up research laboratories they would exhaust the supply of trained personnel.”

Top priorities are the same as elsewhere: electronics and biotechnology. “But Spain should become one of Europe's biggest food producers in the next few years and the biotechnology to be sought is to improve processes in food treatment rather than new pharmaceuticals”.

Attempts are being made to build bridges between the universities, research institutes and industry, but how successful are they? “Efforts beyond the new Law of Science will be required. I have supported giving preference in grants to industry if they are collaborating with the universities, and to those who have joined multinational projects. And the technical universities need more resources”.

Industrial research

Looking for high-tech winners

THE basic research of yesterday is the applied research of today and the product of tomorrow. Most governments already realize that special mechanisms are needed to help speed the transformation of ideas into realities if they are not to be left behind in the international high-technology race. In Spain the need is perhaps more acute than elsewhere, for the banking and industrial communities are conservative and show little taste for risky ventures. So the government has provided the Centre for the Development of Industrial Technology (CDTI).

CDTI was set up back in 1977 but it seems to have only really got cracking after it was set free in 1982. It now has the status of a private company, although its budget and its board of directors are still provided by the government. Its director, Dr Jaime Sodupe, sees it as an “interface between industry and science” but talking more to the industry side as “there are already plenty of institutions to worry about science”. Its activities carry on where CAICYT leaves off; support to help develop a small (or sometimes not so small) company's prototype so that it can enter the marketplace comes a stage after the applied research backed by CAICYT.

Some 600,000 million pesetas are now pumped by CDTI's 50 staff into its 100–200 projects each year. A part of the money is given out in venture capital repayable after two to six years. But the important point is that repayment is link-

ed to commercial success; if the product fails CDTI can say goodbye to its money. Products already close to the market are often financed with an unsecured loan — without a need for guarantees the company's ability to borrow further from the banks is not affected. So far CDTI has had more than 60 per cent of its money returned safely which makes the director wonder if they are being sufficiently daring; similar organizations elsewhere in Europe often obtain only 30 per cent return.

Somewhat unusual is CDTI's role in international affairs. It is responsible for managing Spain's annual contribution to CERN (2,000 million pesetas), the European Space Agency (5,000 million pesetas) and soon the Airbus. And it also tries to find ways to get companies into European programmes. There have been successes. At the Eureka meeting in London in June, 67 projects were involved; 14 of them involving Spain, six of these with Spain as chief partner. CDTI's staff is optimistic, “each month is easier than the last, we don't have to convince managers anymore”, said one of them. One remaining problem is that a big part of CDTI's job is also to collect and distribute information, and make industry aware of the opportunities and the latest world-wide trends (as well as organizing promotion of Spanish industry at fairs and the like). But Spain's embassies abroad are still living in the past — they have no science attaches.

Industry: biotechnology

A sunny spot for entrepreneurs

SPAIN has not been immune from the world-wide rush towards the commercial exploitation of the advances of molecular biology. As elsewhere, the United States has provided the inspiration for scientists to turn entrepreneur — often directly, for the staff of the new biotechnology companies mostly have excellent American accents acquired at top US universities. The risk capital behind them, though, is Spanish: often it is CDTI (opposite) that is playing fairy godmother.

Five small companies are in the market for diagnostic kits using monoclonal antibodies. Among them are the well-established Biokit, now participating in a Eureka project to develop a gonorrhoea diagnosis kit, and the newer Invesgen which among other things is looking for a test for heroin consumption (proving between them that there is money for molecular biology in moral decay).

Biokit, described by a competitor as “a very pushy lot”, can boast that the average age of their staff is just 30 years. They were the first to manufacture diagnostic kits in Spain, with a test they produced for rheumatoid factor back in 1976, just two years after setting up their company with ten people in a “120 square metre space”. World-wide marketing agreements with foreign companies followed for this and for their next product, a syphilis diagnostic. With financial help from CDTI, chlamydia, rubella and hepatitis diagnostics came next. The philosophy of Biokit when aiming at a new product is “we must have contact with everyone in the world with knowledge in that field”. What this means in practice is that despite being based in Barcelona, Biokit scientists go out to the United States knocking on people's doors until they are known and trusted and it is understood that the little company from Spain should be taken seriously.

The future now looks bright. Many of the products under research will be launched in the next year or two and there is hope that sales will reach the 1,000 million peseta mark and Biokit be a world leader in sexual disease diagnostic kits. But still the major input of expertise is likely to be US not Spanish universities. For the distant future (which is more than three years ahead in this business) there is hope from the Eureka project. But the race to develop a quick test for gonorrhoea is not going to be an easy one to win.

Invesgen has not had to work so hard in building links to the United States as its key personnel came from Massachusetts Institute of Technology (MIT). Indeed, Invesgen probably only exists because the government decided back in 1974 to send thirty graduates to MIT to study for their

doctorates. On their return some were involved in a project to set up Spain's own miniature MIT — an idea that eventually proved to be ahead of its time. Two years ago, four of the original MIT students got together with a fifth researcher to found Invesgen, at first within a division of a huge company of the El Corte Inglés group and soon after as an independent company in the suburbs of Madrid.

A start in setting up a sales network came through an agreement to market the Pasteur Institute's products; and a push for new products came with finance from CDTI. The company has grown to employ ten postdoctoral researchers plus technicians and sales staff. Closest to the market is a kit to quantify simply factors related to rheumatic disease (C-reactive protein, streptolysin, and rheumatoid factor). Selling Spanish products to hospitals in Spain is not going to be easy for there is still a feeling that foreign products are better. But the company's research director, Dr Julio Coll Morales, quotes the adage “no-one profits in his own land” and says that progress will come through selling in the United States. Distance from the world's scientific giants does make progress harder; at the most direct level there can be very long waits for reagents, sometimes so long that “people forget why they ordered them by the time they arrive”.

And Dr Coll Morales admits that they are “a little lonely” working in a small group far from the universities. There is still some prejudice against the “fellow from industry”. Hopes are that the new Centre for Biotechnology will make a difference. □

Patent success for pharmaceutical group

FERRER International is an exception to the rule that Spain's companies do not invest in research and technology. The company is not among Spain's 200 biggest enterprises — but when it comes to total research and development expenditure it ranks twelfth. And its researchers boast more patents than anyone else in the private sector. Its business is pharmaceuticals.

Change in the company's thinking — a determination to make new rather than derivative products — came in the 1970s alongside the nation's political liberalization. An effort was made to bring research up to best European levels and now over 10 per cent of Ferrer's 1,000 employees work in its Barcelona research centre.

For the centre's director, there is no choice for Spain but to try to increase its technological level, “it's either technology or colonization”, he says. But the path ahead is by no means easy; there are plenty

Ports safe in a storm

PROFESSOR Felipe Martínez and Dr Manuel Pastor of the Research Centre for Public Works (CEDEX) stand in front of Bilbao harbour. Not the real thing, but a massive model set up in a gigantic building complete with wave-making machines. Under the same roof are the harbours of Valencia, Las Palmas, and part of Barcelona beach needing restoring before the Olympics. The Centre for the Study of Ports and Coasts, where Professor Martínez is director, looks at all kinds of problems in coastal



defence and harbour construction: particularly important as Spain has one of the world's largest fishing fleets, much of it operating from small ports. At Bilbao the problems are for big ships: a peculiar resonance in the harbour at times amplifies long-period waves (of four minute period and unknown origin) and makes unloading of ships at one quay difficult. Dr Pastor's finite element model and the physical model itself both produce the same effect and suggest a cure — opening an extra narrow channel between two basins. □

of giants in the pharmaceutical business and Ferrer International is just medium-sized, with annual sales worth some 15,000 million pesetas. To succeed, the company will have to rely on “agility” and sticking to fields where they have good ideas, mainly analgesics, antidepressants, Ca^{2+} antagonists, H_2 histamine receptor blockers and antifungal drugs.

Their approach is conventional: studying physiology and the structure of chemicals known to have physiological effects modifying and testing them and improving routes of chemical synthesis. Collaborations with universities are common: “universities are changing”, says the director, “but we must be realistic. It is difficult for them. They often do not have the best apparatus and they must think of how many publications they can produce. Industrial interest is not yet sufficient to resolve their problems.” □

The Basque Country

A mountain of problems to tackle

THE Basque Country suffered from cruel neglect under the Franco regime. Its heartland, the provinces of Guipuzcoa and Biscay (which contains the town of Guernica), had fought on the side of the Republic during the civil war and been granted autonomy by it. With Franco's victory autonomy ended; the region was 'punished' and use of the Basque language was banned. Despite the region's enormous industrial importance the central government refused to establish a university or research institute there — just a school of engineering. Now the technological backup to help replace the region's declining heavy industry is just being developed. And out of hardship there has grown a unique Basque response in indigenous cooperatives that link industries, research institutes and a bank.

Parts of Bilbao, where heavy industry was most concentrated, are palpably dying. The town grew rich in the First World War when Spain was neutral and both sides were desperate to buy its iron and steel. Several of its biggest works are now bankrupt but the city, trapped in a narrow valley, still spends half its time blanketed in smog. Signs of decay are everywhere; the unemployed crowd street corners and the walls are covered with slogans, both political (in Basque) and the more desperate "*Heroína Assassasina*". The worst blows, occasioned by entry into the EEC following on a world recession, are thought to be over. The contraction in the steel- and ship-yards is coming to a halt but a large number of small and medium-sized manufacturing businesses remain that have never had the money to invest in research and whose equipment is out of date. And in a story familiar from Sunderland to Pittsburgh, the race is on to add value: control and automation, computer-assisted design and manufacturing, robotics and flexible manufacturing are the key words.

There is only one area that is taboo: nuclear power. Construction of a nuclear power plant was begun but abandoned after the project manager, and then his successor, were assassinated by Basque Nationalists. Although nationalist sentiments have lessened with regional autonomy, terrorism has still to be taken very seriously.

The regional government is doing what it can for science and technology within its limited means. It has taken over responsibility for several research centres intended to raise industry's capabilities. Typical of them is Labeine in Bilbao. Set up in 1955, when research was more a "desire than a reality", it has expanded at a cracking pace (40 per cent a year) since the local government gained responsibility in 1980.

Its budget is now 600 million pesetas (£3 million), still peanuts by international standards; and its staff 130, half of them scientists. The intention is not to do basic research but help industry; both at the level of testing and pollution control and the introduction of high technology. Among their projects are those for developing an expert system to diagnose faults in electric power stations and to establish a unit for the supply of customized integrated circuits.

A new university has been established, scattered over the region but with most of its departments in Bilbao and San Sebastian. It began life in 1968 as the University of Bilbao; addition of other centres turned it into the University of the Basque Country, now with a little over 40,000 students. The quality of the university remains very uneven and it proves hard to recruit good people in those areas, like computing and engineering, where they are most needed, for industry can always offer better pay. A massive shortfall of qualified people in these areas is forecast for the end of the century, not so much through lack of students but through lack of competent teachers.

Basic science is being provided for in scholarship schemes that are unique to the Basque region, but which the central government could do well to emulate. One hundred and fifty predoctoral and twenty postdoctoral grants are being given (held for two years in Spain, or three years in a foreign country) to help raise young researchers' standards. About half the grant holders stay in the country; for the rest France, the United Kingdom and then the United States are favourite destinations (and cost the government only a little over

a million pesetas a year). Attempts are also being made to bring the university up to international level by providing money for re-equipping laboratories. Four hundred million pesetas were spent in this way last year, but applications were for four times that amount and the search is on for new sources of funds. Industry is more directly supported through spending 1,000 million pesetas a year on projects in the regions' five industrial research centres (including Labeine and Ikerlan described here) to develop 'generic technologies' to the level where industry can come in for specific applications.

All this support comes from local taxes, quite separately from payments made to the central government for 'national' services — which still includes science and technology. Some of the money comes back in grants from CAICYT but 70 per cent of the nation's research institutes are found in Madrid. Not surprisingly there are local politicians who strongly feel that the contribution for the science budget should be handed back to the Basque government to spend as it wishes in its own area.

The difference of opinion stems from the failure of the constitution to be crystal clear on the dividing line between the competence of the autonomous regions and the central government. Does the power of the government to coordinate research mean that it also says how the money is to be spent? The new law of science, which increases central control, raises hackles in the Basque Country and many declare it unconstitutional. More rumblings of discontent from them (and from Catalan politicians) are bound to follow but it would take a narrowly won election, with nationalist parties holding the balance, to persuade the central government, whatever the constitution may say. □



Mondragon, amid mountains in the heart of the Basque Country, is packed with industrial cooperatives.

The Basque Country

Cooperatives provide a solution

ANYONE who thinks that cooperatives, in which workers own their workplaces, could never be more than an idealist's dream should visit Mondragon in the Basque Country. Britain's Minister of Employment, Kenneth Clark, made the effort last month and apparently came away impressed. Mondragon's industrial cooperatives now have 19,000 employees (or employers, depending on which way you look at it) and their export sales are booming. Alongside them are agricultural, education, consumer, housing and services cooperatives which can provide everything from technical education to part-time jobs for women with children. And the industrial cooperatives *en masse* have felt the need to raise their technological standards sufficient to establish Ikerlan — a unique cooperative research institute.

The cooperative movement grew in the special conditions of the Basque Country from the vision of one extraordinary man, Father Don Jose Maria Arizmendiarieta. During the civil war he worked as a Basque Army journalist and at its end he narrowly escaped execution. He came to Mondragon, a small town in the mountains south of Bilbao, and in 1943 set up a small technical training school with the aim of aiding the reconstruction of his country. Those enrolled not only studied but raised money from the community to keep the school going. As it turned out, the graduates of the school were later to become the managers and technicians of the cooperatives. The first was established in 1956, after a vain attempt to persuade local industry to allow worker participation in ownership. It consisted of just five men making paraffin heaters. Within a year there were five more cooperatives. They were a success; perhaps partly because the Basque Country has a long history of communal labour on its farms. Then the problem became how to raise capital for further expansion. The answer: set up a cooperative bank. More than twenty years later the Caja Laboral Popular (the working people's savings bank) has over 120,000 million pesetas (£600 million) on deposit and provides financial management, export advice and banking services to the region's 160 plus cooperatives. Its funds have come from the local community, who knew their money would be invested in cooperatives, and from the profits of the cooperatives themselves.

Each 'primary' cooperative is based on the simple principle that all who work there must be shareholders, and no one can be a shareholder who does not work there: the "indissoluble combination of worker and employer". The board of directors is elected and the highest-paid can

earn no more than three times the salary of the lowest paid. The bank, is a 'secondary' cooperative, it provides services to all cooperatives, and has its management elected both by its own 1,000-plus workers and by the other cooperatives. The research institute, Ikerlan, set up in 1977, is another secondary cooperative.

The word 'cooperative' might conjure up an image of low-tech — of bearded idealists making handicrafts in pastoral surroundings — but that is not what Ikerlan is about (although the pastoral surroundings are there). Its aim is to develop the technology to make the cooperatives internationally competitive. Its laboratories are packed with computers, robots, experimental flexible manufacturing systems, and even a miniature 'robotized transportation vehicle'. It differs little from any small advanced technology institute from Tokyo to San Francisco.

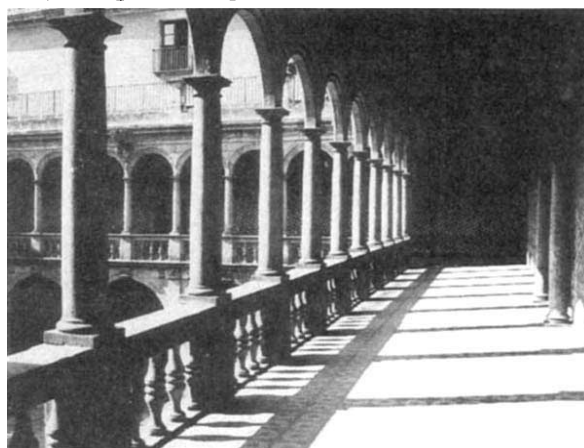
Technological support is given to every cooperative (and other business) that solicits its services. Well over a hundred

projects aimed at improving someone's product have been completed in its 12-year history. The research always involves members of the organization that requested it in order that training and technology transfer can be combined with product development — a philosophy that "was at first hard to get over to the government" which now provides for part of the running costs of Ikerlan. It is recognized that specific product development alone will not suffice to keep local industry competitive; half of Ikerlan's 70 scientists' time is now spent on generic research — basic work likely to later find specific application. In the computer sciences, for example, there are research programmes on geometric modelling, automatic programming of numerically-controlled machines and interactive computer-aided drawing. Just as elsewhere the aim is to provide a computer environment which helps an engineer to go from drawings for a specific product generated with the help of a computer to instructions for its computer-controlled creation. It is to the implementation of such research that small industries, producing say complex gears, look for their long-term health. □

Can there be a real 'Catalan science'

THE seventeenth century cloister shown here provides the entrance way to the Institute of Catalan Studies, maintained in rooms originally built as the house of convalescence to the fourteenth century hospital alongside. The institute is not as old as that, having been set up in 1907 on the crest

the arrival of the Second Republic (1931–39) it flourished, and then when Franco came to power it had to continue its work of publishing scientific books in Catalan underground. Since 1976 things have returned to normality and the institute has returned to its original headquarters.



of a revival in Catalan culture to aid "high scientific research" in the region.

Within a few years the institute had a structure rather like that of the old academies, with four sections: philology (mainly the study of the Catalan language), history and archaeology, science, and philosophy and social science. But its fortunes were thereafter to change with the political seasons, for any organization that might foster regional autonomy was suspect. Under the military dictatorship of Primo de Rivera (1923–30) it was suppressed, with

A large computer now hums away under the gothic arches in the basement helping to prepare a new dictionary of contemporary Catalan. For scientists, its principal importance is that all the local scientific societies are affiliated to it and rely upon it to help organize conferences and prepare publications. It has set up a Centre for Mathematical Research in the Autonomous University of Barcelona and is running joint projects; with the University of St Louis on the

microseismicity of the Pyrenees, and with a Europe-wide group on the mapping of bryophytes.

The institute's president, Professor Enric Casassas i Simo, would like to see the institute expand further into a true interdisciplinary research centre. But there are others who see the institute's involvement in science as an anachronism. They argue that there is no place for "Catalan science" in the international age and particularly deride attempts to find Catalan equivalents for new scientific terms. □

Catalonia

Plenty of energy in Barcelona

THE really big news in Barcelona of course is not science but the Olympics. The choice of Barcelona as the site for the Olympic Games in 1992 sent Catalonia wild and will do much to heighten regional pride. It will also do something for science: a new biomedical institute is certain to be created for toxicological tests on the athletes and will live on after the games. And an Olympic village is going to be necessary and might afterwards provide much needed halls of residence for the universities (pessimists say the village will simply swallow up funds previously earmarked for the universities).

Like the Basque Country, Catalonia has a long history of independence and its own language and culture. But emotional appeals for independence are rarely heard, perhaps because Catalonia is recognizably prosperous and contains many of Spain's high-tech industries (as well as the Costa Brava). And it now has three excellent universities. In the Franco era the region was neglected. Despite the intellectual eminence of Barcelona, Spain's most cosmopolitan city with a population second only to Madrid, just four of CSIC's 78 institutes were located here: those for fisheries, humanities, geology and the Centre for Research and Development. New institutes are now springing up for microelectronics, materials science and economic analysis.

The region has its own commission for science and technology (CIRIT) and more than a few academics who would like to

see Catalonia fully responsible for disbursing its own contribution to the central science budget. But until the unlikely day when Madrid relinquishes this power, CIRIT will have to make do with funds the Catalan government can raise to supplement central government allocations. All agree it cannot raise enough. Around 100 million pesetas goes each year in sending people abroad for training, and a further 200 million in research grants. Other funds are available for research in specific sectors, such as agriculture, but overall there is insufficient money for any large, specific project.

Of the CSIC institutes only the Centre for Research and Development is of any real size. As elsewhere, the institute has been through a painful process of reorganization. Just last year the five-storied building contained five institutes, with five directors, five libraries, and five administrators, all run separately. Now one institute, it brings together all the research groups in Barcelona working on pharmacology, organic chemistry and molecular biology. Changing the structure was clearly not easy: a bitter dispute over the election of laboratory heads has left one floor of the building with a crowd of researchers at one of its ends and two empty laboratories stripped of equipment, with one professor, formerly an institute chief, at the other. It is now up to the courts to decide whether the elections were held properly, and what belongs to who.

Overall director of the institute is Pro-



Sir Alexander Fleming is much fêted in Spain with streets or squares named after him in several cities. Above is a famous statue situated in a small square at the edge of the Barcelona red-light district. It is said that its inhabitants often leave flowers here in gratitude for his discovery of penicillin.

fessor Joan Albaiges Riera. He sees the key advantage of the reorganization as creating groups of above critical size (at least 3 or 4 people in each, plus research students and technicians) and bringing related groups together to promote interaction. His speciality is environmental chemistry and entry into the EEC has greatly helped extend international connections and thus raise the level of research. The institute will be taking part in both the EUROTRAC and EUROMAR programmes to monitor air and sea pollution. □

An education that pays for itself

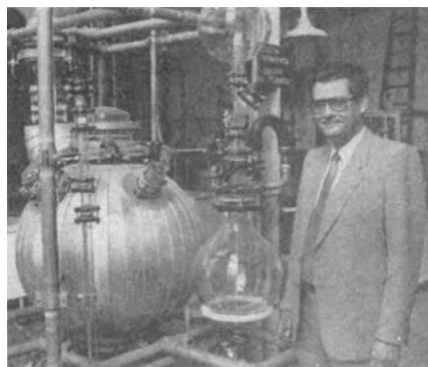
In an era of rapid change it is perhaps no surprise to find the occasional survivor from another age. Such is the Instituto Químico de Sarria, set up in 1916, which can be summed as a "private Jesuit university with only one faculty — that for chemical engineering". State schools of engineering turned into polytechnics in the 1970s, and are now, if anything, more prestigious than the universities. Sarria continues as a single-subject university — but that does not mean it has lost prestige.

The institute, in an old suburb at the edge of Barcelona, shows its origins. At its centre are high-ceilinged laboratories, elegantly tiled in the fashion of the 1920s; across the corridor is a dimly-lit chapel entered by carved wooden doors.

But the institute is no fossil — courses are now well tailored to the needs of industry. A survey of its graduates finds 31 per cent of them working as directors in industry and another 42 per cent as technical executives. The institute's director, Dr Miguel Gassiot Matas, reckons there are two to three jobs available in industry for

each graduate. The ability to guarantee their students a job seems a hallmark of all the Jesuit universities.

Part of the reason for the institute's success is its belief that it is not possible to teach well without doing research; a view



The director of the Instituto Químico de Sarria, Dr Miguel Gassiot Matas, in his teaching laboratory. The apparatus is not all just for demonstrating industrial processes; the students also help manufacture pharmaceutical caffeine which adds to the institute's income.

that has only recently become fashionable in the state universities. Indeed the institute has to do research in order to survive; its costs are not met by the state and even with the high fees (around 300,000 pesetas, £1,500 a year) paid by its 600 students, it cannot make ends meet.

Almost a third of the institute's income stems from contract work, performed for industry and government. Much of the work is analysis of food, chemicals and pharmaceuticals but forensic science is also important. It was to the institute that the government turned during the outbreak of toxic oil syndrome, although it was a laboratory in London that discovered anilides in the oil a few days before them.

What of the future? The institute's director is well aware that each branch of science grows ever-more dependent on its neighbours; he looks to construct curricula connected with different fields to create "a small polytechnic school centred around chemical engineering". And with a look back to the political past he says, "an option to the state system is necessary; it is the most important guarantee of liberty". □

Portugal: political change

Dictatorship out with a whimper

JUST twelve years ago Portugal lacked a democratic government and was fighting wars in its African colonies of Mozambique, Guinea and Angola which had dragged on for a decade. A quarter of its male population of military age was in the armed forces. Now democracy is firmly established and of the colonies only Macao remains (until China takes it back).

Although change has been rapid the recent relationship with the colonies still colours life more than in any other European nation. Almost a twelfth of the population has lived in Africa for when the military struggle was finally abandoned some 750,000 settlers came home. But as the Empire that stretched all the way east to Timor is becoming a memory, Portugal is turning to enthusiastically embrace Europe.

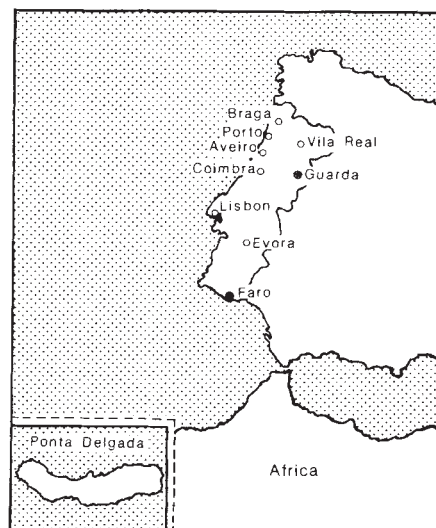
The series of political changes that led to the restoration of democracy, beginning with the 'revolution' in 1974, were triggered by weariness with the colonial wars. The dictator Antonio Salazar, who had ruled Portugal for 40 years, was already long out of office, having been paralysed by a stroke in 1968. But his successor, Marcelo Caetano, proved unable to modernize the country. Elections were held in 1969 but barely 20 per cent of the population were permitted to vote and opposition parties strictly supervised. The unsurprising result was a landslide victory for the ruling party.

Just how fragile the real power of the

government had become was demonstrated on 25 April 1974. A group of army officers led a coup in which, with only a few shots and no casualties, over 40 years of dictatorial rule was brought to an end. Under the new rulers of the Armed Forces Movement power was placed in the hands of the Junta of National Salvation. At first, it took a moderate line but within a few months it lurched leftwards.

A Supreme Revolutionary Council was formed and began to implement radical policies whose effects are still apparent in Portuguese economic life. Large parts of the economy were nationalized, including banking, insurance, shipbuilding and transportation. In the south of the country, in the Alentejo, large private farms were seized by their workers and divided up. And the colonies were granted independence. But a massive turnout for the election of 1975 (to choose an Assembly to draw up a new constitution) gave support to the socialists, not the communists.

An unstable period followed, with attacks by communist sympathizers on the parliament, culminating in an attempt at a coup by left-wing military officers in 1975. Its failure proved the beginning of true democracy in Portugal. In the elections that followed the approval of the new constitution in 1976, the socialists formed the first freely-elected government. Since then the constitution has been revised giving Portugal an elected President, who appoints a Prime Minister, and



Portugal's twelve state universities: Azores (at Ponta Delgada), Algarve (at Faro), Aveiro, Coimbra, Évora, Lisboa (Lisbon), Nova de Lisboa, Técnica de Lisboa, Minho (at Braga), Porto, Beira Interior (at Guarda) and Trás-os-Montes e Alto Douro (at Vila Real). There are also a number of small private universities and a large Catholic University in Lisbon. The universities at Vila Real, Evora, and Faro are very new and are intended to act as poles of development in the inner regions of the country. Here conditions are backward and there is enormous scope for improvement in agriculture. Portugal has an illiteracy rate of 19 per cent and is still in desperate need of training schemes in basic technology. 90,000 students are now in higher education. A network of polytechnics is being set up to complement the universities.

an elected unicameral legislature, the Assembly of the Republic.

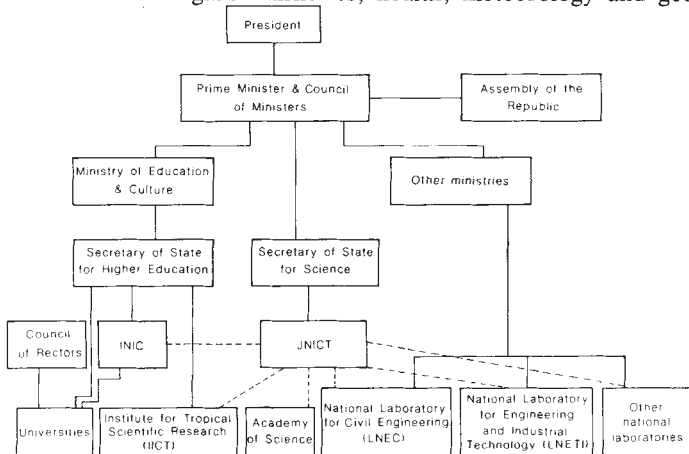
To date no party has ever been able to gain an absolute majority and changes of government have been embarrassingly frequent. The Social Democrats have now overtaken the socialists as the largest group, although they are in turn easily outnumbered by the combined opposition (or by just the socialists and communists). There has been a general trend rightwards (the Supreme Revolutionary Council, which persisted as an advisory board was abolished in 1982) but the massive state holdings acquired earlier remain. Their status is highly controversial: they have performed poorly and the banks have proved unable to match the performance of smaller foreign entrants.

The lands seized in the Alentejo are gradually being returned to private hands as mistaken agricultural policies (including the replacement of the traditional cork trees by wheat) have failed. But the strength of the left prevents any rapid reversal of former policies. Key issues now are the need to raise the general educational standards (the illiteracy rate is the highest in Europe) and to improve agricultural practices in the poorer regions as barriers protecting Portugal from full exposure to EEC competition are gradually removed.

Pattern of research planning explained

THE Junta Nacional Investigação Científica e Tecnologia (JNICT) is growing into the major body for the planning, promotion and coordination of Portugal's scientific and technological research. The overall control of JNICT goes to the Secretary of State for Science who operates under the jurisdiction of the Minister of Planning. Much of the research within the universities is carried out in centres that belong to the Instituto Nacional de Investigação Científica (INIC), a funding agency operating under the Ministry of Education and Culture. Besides the twelve universities, one remarkable research institute, the Instituto de Investigação Científica

Tropical (IICT) is controlled by the education ministry. Other ministries control research institutes for agriculture, fisheries, health, meteorology and geo-



physics, and hydrographics. Best known are the Laboratório Nacional de Engenharia e Tecnologia Industrial (LNETI) and the Laboratório Nacional de Engenharia Civil (LNEC). □

Science policy

A support system in evolution

PORTUGAL is on the verge of taking scientific research seriously. Somewhere in its research organization exist bodies to monitor international trends and identify national priorities, evaluate fairly project proposals, coordinate different sectors, and encourage interaction between industry and academia — all the things a modern nation needs. And it has a supply of well-trained researchers. The problem is that the various parts are not all the right size.

Portuguese scientists are themselves well aware of the problems — even enthusiastic about tackling them. Change has begun. In the scenario now developing two organizations will certainly be important: the National Board for Scientific and Technological Research (JNICT) and the National Institute for Scientific Research (INIC). The former's star is on the rise while the latter, even by the admission of those who run it, could profitably be reformed in a more sophisticated organization.

JNICT

As the Minister of Education, Professor João de Deus Pinheiro, put it, "JNICT was the turning point". When it was set up in 1967 it was the first organization to evaluate projects using independent referees, including foreigners, and to follow up and assess how well projects were running. Now JNICT appears as the one national body planning, coordinating and funding scientific and technological research in all sectors of activity. But it is still not the real scientific power in the land it could be, despite its 20 year history.

Two things have held JNICT back: frequent changes in its position in the administration (it is now in the Ministry of Planning, under a Secretary of State for Scientific Research, but this is its sixth resting place) and, more important, the failure of the government to give it sufficient funds. The total science budget now runs close to 13,000 million escudos (65 per cent of which is government expenditure, most of the remainder coming from industry) but last year JNICT received only 200 million escudos (1.5 per cent of the total).

The future looks brighter. This year the budget rose to 1,000 million escudos and next year it should go 2,500 million escudos (19.2 per cent of the total), for the first time putting it ahead both of INIC and the largest public research institute, the National Laboratory of Engineering and Industrial Technology (LNETI).

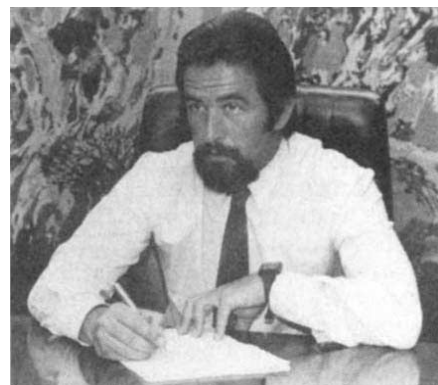
JNICT's grant can flow anywhere within the research and development system — to universities and their research centres, to state laboratories, to industry, and to

non-profit research organizations. It is thus well placed to ensure that national priorities are met and gaps filled. In the words of the JNICT president, Professor José Mariano Gago, "the strongest indication of a trend to increase overall coordination is the amount of money being put here." A more powerful JNICT with what Professor Pinheiro called a dual policy of backing "both the best researchers and the areas that give an advantage to Portugal" could be just what the country needs. With its bigger budget its plans will grow. Next year the largest slice will go on contract research, concentrated in four priority areas: material science, marine science, biotechnology and electronics. Part of the remainder will go to encouraging industrial applications of research but a sizeable portion will be reserved for "encouraging excellence": ten to fifteen groups of real international standing will be supported whatever their field of study might be.

One problem that JNICT is trying to cure is that of "small-scale groups who can never achieve good results in time". Attempts are being made to create new networks of researchers. Membership of CERN is providing some opportunities for at first the annual contribution will be spent in Portugal building up resources. A Lisbon group specializing in microcomputers and optical communication will link up to a Coimbra group with expertise in nuclear detectors and instrumentation to form a new 'institute'. The share out of funds is entirely decided by an international committee. A similar scheme will create a new institute of biotechnology.

INIC

For the universities the Ministry of Education's own grant body, INIC, is of prime importance in supporting research. That support does not, however, go directly to the universities but mostly to those in INIC's own 126 research centres within the universities. The 3,500 researchers in the centres are almost all university professors so their salaries come from a separate source but they have fewer teaching and administrative duties that staff outside the centres. That leaves both INIC and the universities in a curious position; INIC has a large number of quite small research centres whose staff's promotion is decided by the university, while the universities have no real control over the research activities of their own staff. Worse still is that research centres covering the full range of disciplines have to be maintained and that means INIC is under obligation to continue funding them all by block grants. Projects assessment is included in the complicated formula used to calculate



Portugal's Minister of Education, Professor João de Deus Pinheiro has been in his job two years "an almost unimaginably long time" given the frequency with which governments change. Total expenditure on research he sees as low although it is probably double the official figure of 0.4 per cent of GNP because "salaries are not included". But change is going to involve more than just increasing budgets. "For the fifty years of authoritarian government a good student was one who could repeat the text book or the teacher's words. Creativity and criticism were words absent from our education system" he says. A massive review of the education system is now under way in which priority will go to improving science teaching in the schools through teacher retraining. To improve the research system help is being sought from the World Bank for institutes in biotechnology and agriculture, water resources, and materials. An institute for marine science is also planned. Improving links with industry is a priority. Inter-ministerial coordination is still poor and relations with the Ministry of Industry have "not always been good".

Young Portuguese, the minister says, might learn something from thinking afresh about history. Rather than viewing the fifteenth century voyages as an outcome of the desire to spread Christianity they can be seen as stemming from wise technological policy. Portugal's breakthroughs in ship design and organization created the equivalent of "today's space rockets" and made exploration possible. □

each centre's annual budget but there is no strong sense of competition for research funds and, in any case, INIC's referees are not noted for their critical powers.

Originally, this system made sense in that central distribution helped overcome the universities' own problems in sharing out their budgets. But as the universities have matured and a Law of University Autonomy looms on the horizon there seems little point in the present system. INIC officials are themselves proposing change. Many of the centres, they say, could be handed over to the universities and instead some twenty or so larger interdisciplinary institutes be created. The new organizations springing up to serve high-energy physics and biotechnology could provide a model. The problem is that changes of this kind have been suggested for some years, without much happening. The President of INIC, Alberto José Correia Ralha, a very patient man, says "sometimes even I think I am going to get angry". □

Tropical research

Everything under the Sun

THE only way to describe Lisbon's remarkable Instituto de Investigação Científica Tropical is a 'horizontal institute'. Rather than specializing in one field, it covers almost everything provided it can be related to the tropics. Geology, archaeology, zoology, ethnology, geodesy, veterinary medicine, history, ancient maps, soil science and anthropology are studied among other things, and there is a botanical garden.

There is a historical explanation for the institute's structure. It began life just over a hundred years ago as a Commission for Cartography to map the new colonies and, as the colonies grew, so it expanded to offer them a complete range of scientific services. And if it had not been for the British, the colonies would probably have been a lot bigger. Maps produced early this century show everything between Angola and Mozambique painted the 'rose colour' of Portuguese territory. But the British sternly warned that their north-south expansion was not going to be halted by a Portuguese east-west connection. The resulting crisis and threats that gunboats were to be dispatched brought down the government of the day.

After the colonies were given up, twelve years ago, there was a series of upheavals (along with changes of name from 'Junta' to 'Laboratorio' to 'Instituto') leaving the institute in the care of the Ministry of Education with the aim of aiding scientific research in the tropics (in 40 countries, not just the former colonies) and training specialized personnel. The government budget is 800 million escudos a year but further help comes from international organizations; the United Nations, UNESCO and the African Development Bank among them. The different specialities are not housed under one

roof but in independent centres and each is connected to relevant university departments: the directors are also university professors.

For the international community perhaps the most important centre is that for the Investigação das Ferrugens do Cafeeiro (CIFC), a world centre for studying coffee rusts — fungal diseases. Coffee may be only a drink but it is of tremendous economic importance: second only to petroleum in the value of its world trade.

Rusts are important because of what they could do to the industry. The major commercial species of coffee is *Coffea arabica*. But *arabica* is highly susceptible to coffee leaf rust, which quickly destroys the plant. In the first major outbreak, towards the end of the nineteenth century, Ceylon's coffee industry was utterly destroyed. Little natural resistance has been found among varieties of *arabica*. And to make matters worse, it appears that all the coffee plantations in Central and South America, where 65 per cent of the world supply is grown, were established from the progeny of a single plant, taken from Java to the botanical garden in Amsterdam in 1706.

Angola, Portugal's biggest African colony, was the world's third-greatest coffee producer. But the coffee rust research centre in 1955 owes less to this link than to US worries about what an outbreak of coffee rust could do to the economic stability of Latin American countries dependent on coffee crops. In 1952 two US scientists set out on a world-wide tour to look at the disease and, after encountering a Portuguese researcher, decided Lisbon would be the ideal place for a centre — no coffee was grown in Portugal and rusts from around the world could be studied safely there.

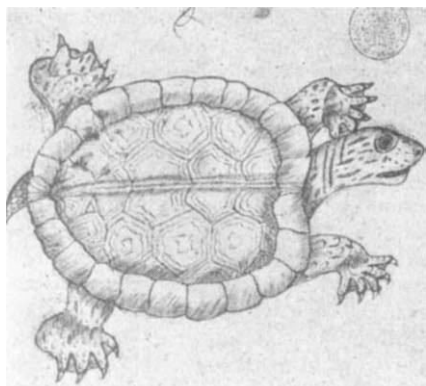


Coffee is big business and a fungus, coffee rust, a big threat to it. Here just a few of the tens of thousands of coffee plants that pass before the Centro de Investigação das Ferrugens do Cafeeiro director C.J. Rodrigues Jr for testing each year.

In those days, the fungus had not reached the Americas. But in 1970 the first outbreak occurred in Brazil, and within a few years the rust had spread all over the country. Spores are probably carried by the wind for they have been found 2,000 metres up in the air. By 1983 rust had crossed the Andes into Columbia and also appeared in Central America.

In the intervening years the centre had accumulated some 33 different races of the fungus and a vast collection of coffee varieties that had enabled them to classify plants according to their resistance spectrum. With this standardization done they could begin work on the genetics of resistance. And they had discovered in another of their colonies, Portuguese Timor, a hybrid between *C. arabica* and *C. canephora* that seemed totally resistant to almost all fungal races. Furthermore, like *arabica* it was tetraploid and self-fertile. Luckily the resistance genes were dominant. A programme of crossing and back-crossing and resistance testing then began to introduce them into an *arabica* background.

Altogether nine genes for rust resistance have now been identified. The centre is cooperating with researchers all around the world to help develop resistant strains, with the 'Timor hybrid' still at the centre. Breeders send tens of thousands of coffee plants by air every year for resistance testing in the centre's greenhouses — for many growing regions the choice may soon become resistant strains or crippling bills for fungicides. Plans are now afoot to use cryogenic techniques to preserve spores and to develop tissue culture methods so that known lines of coffee plants can be propagated quickly.



GIVEN its long history and colonial links it is not surprising that the Institute for Tropical Science has amassed some treasures. Above is an illustration showing a turtle from one of the earliest works on the

natural history of Brazil, the *História dos Animais e Arvores do Maranhão*, by Father Cristóvão de Lisboa who was resident in Maranhão in 1627. The monograph is on display at the Archives of Overseas History which now houses the records of the old Overseas Territories Office as well as other collections.

Another treasure house, too little known outside Portugal, is in the Centre for Ethnology. For years it has put its collections on public view only rarely but now it is rigorously pursuing an open policy. An astonishing collection of African sculpture is now on view (all, alas, the product of cultures already destroyed) and in the basement store-room glass fronts are being added to crowded storage shelves so that the public will be able to see everything the museum has.

Portugal's university research

Adding a little more flexibility

RESEARCH funds from both INIC and JNICT are available to university researchers. But according to a group of scientists drawn from the New University of Lisbon's Faculty of Science and Engineering, the total does not necessarily add up to much: "Just a little from each, only occasionally a jackpot". In fact, in a good year there are around 120 million escudos (£600,000) to support the research of the 70 staff, including money from industrial sources.

This university is now trying, with a lead provided by Professor Leopoldo Guimarães, to create an alternative in UNIN-OVA, an institute bringing together the universities and industry. Industry has not been totally neglected in the past. The faculty's Department of Environmental Sciences real research is paid for by contracts — a biomass-from-wine production effluents project paid for by a German company, and environmental assessments projects from the Ministry of Industry and the local municipality. UNIN-OVA — the Institute for the Development of New Technology — is intended to be on a different scale. Some 16 companies, including ICI, Sperry and Digital, are taking shares in the institute. The idea is that the involvement of Portugal's biggest (and often foreign) companies in electronics, computers and biotechnology will power research in those areas as well as providing funds. The institute is not intended to perform research directly for

the companies involved.

A similar organization has already proved a success at the Technical University of Lisbon, and the vice-rector there, Professor Anotónio Simões Lopes, is particularly proud of it. It also has the Institute for Systems and Computer Engineering set up jointly by the university, the University of Porto and the three telecommunication operators. Just six years old it already provides for 300 researchers.

Such institutes will add new flexibility to the university system. It has not proved easy in the past to add numbers where they are most needed, partly because PhD students are virtually guaranteed jobs in the same university so there are limits on how many of them can be taken in.

Despite these difficulties it would be wholly wrong to think of university staff as demoralized. Up to the late 1970s many people were sent abroad for doctoral training, particularly to Britain, and have provided Portugal with a distinctive core of enthusiastic young (under 40) researchers. Every well-established university has a group of international standing. Recent successes (in international collaborations) include the isolation of LAV II virus (with Luc Montagnier of the Institute Pasteur) and the controversial proof of Poincaré's conjecture (with Colin Rourke of the University of Warwick). And Coimbra's nuclear detector group has even attracted a doctoral student from Japan. □

Oiling the wheels

"It's difficult to imagine what Portuguese intellectual life would be without the Gulbenkian Foundation" said Ernesto Veiga de Oliveira, ex-director of Lisbon's Museum of Ethnology; a comment that well summed up what many other researchers had said about the foundation.

The foundation gives away more than seven thousand million escudos each year and even though science is the least of its beneficiaries (receiving 12.6 per cent of the total; after education, arts and social welfare) its impact is clearly felt in the research community. But just why is the Gulbenkian Foundation based in Portugal?

It is a complicated tale. Calouste Gulbenkian was born in Istanbul but took British nationality and was a graduate of Imperial College, London. There he gained a knowledge of mining engineering that helped him realize the potential importance of Middle East oil. By the time he died in 1955 he owned 5 per cent of Iraq's oil reserves. In between he made his home in Paris but, in 1942, after the German invasion, he moved to Portugal, which re-

mained neutral (although favouring the Allies) during the Second World War. He hired a hotel (now the Sheraton), brought in a couple of Rolls-Royces and, finding Portugal to his taste, continued to live there after the war. Along the way he collected the paintings, furniture, ceramics and silver that now form the collections of Lisbon's biggest museums.

The foundation played a key role in Portugal's scientific development. Starting in the mid-1960s, when the universities were at a low ebb, hundreds of scientists were given grants to study abroad. Many now occupy important positions in the universities and government. An institute of science was also established but as the universities developed, the need for it has lessened and it is gradually being closed down. Fields of strategic importance (informatics is now the centre of attention), are helped with special equipment grants. Translations of technical works are made so that inexpensive editions can be published alongside books from Portuguese authors. And links with other Portuguese-speaking countries are kept up through special cooperation agreements. □

Industry

No shift from ivory towers

PROFESSOR Clemente Pedro Nunes, Director General of Higher Education, is new to his job and bursting with energy. He is a chemical engineer, with a doctorate from the University of Birmingham, and plenty of experience of working in industry. The great problem he wants to tackle is the failure of the universities and industry to interact. One statistic sums it up: there are now over 2,000 holders of doctoral degrees in Portugal but less than 10 are working in industry. And there are no signs of a trend to move into the industrial sectors, except in computer software.

Much of the blame must lie with industry for failing to invest in research and development. "Most of all what is needed", says Professor Nunes, "is managers in industry who can discuss science and technology; even in big companies there are very very few." But the universities are also at fault as the progression from a masters degree through a doctoral degree to lifetime tenure is almost assured. There is no pressure to leave and "naturally every professor wants new



Professor Nunes wants more scientists in industry. blood in the department and knows every trick in the book to force extra vacancies. The bill, of course, comes to us (the Ministry of Education). And university salaries are not much different from those in industry."

Fierce words, perhaps, but how can industry be persuaded that research and development is in its best interests? Tax incentives and perhaps "a sabbatical year in industry paid by the state" suggests Professor Nunes. Part of the problem is that "industry is undercapitalized. Nationalization in the mid 1970s broke the back of many structures and they have not yet recovered". Encouraging signs are coming from the joint university-industry research institutes set up within universities. Polytechnic universities also hold out promise for raising industry and agriculture's level of sophistication. There are now 16 in existence, but not all are complete and last year they could take in only 1,000 students. But this year is to be the "big year for their launch". □

Superfluid ^3He in rotation

V. P. Mineev*, M. M. Salomaa† & O. V. Lounasmaa†

* L. D. Landau Institute for Theoretical Physics, USSR Academy of Sciences, 117334, Moscow, USSR

† Low Temperature Laboratory, Helsinki University of Technology, SF-02150 Espoo, Finland

Experiments on vortices in rotating superfluid ^3He have revealed phenomena which have not been observed in any other quantum liquid and have implications in fields beyond low-temperature physics. In $^3\text{He-A}$, vorticity may be supported by a continuous winding of the order parameter which gives rise to the observed continuous coreless vortices with two flow quanta. Exotic vortices with half-integer circulation quanta may also exist in this phase. In $^3\text{He-B}$, an abrupt change in the vortex-core structure has been explained theoretically as a topological phase transition involving half-quantum vortices. NMR experiments on rotating $^3\text{He-B}$ have revealed that the cores of vortices observed possess spontaneously broken parity, manifested by their ferromagnetic superfluid cores.

QUANTIZED vortices in rotating superfluid ^3He display novel structures that make them interesting objects in condensed-matter physics. This can be attributed to their anisotropic nature and rich internal structure of ordering: unlike superfluid ^4He (He II), the superfluid phases of liquid ^3He are not only superfluids but also liquid crystals, magnets, and spin liquid crystals; a few, or all of these different aspects of ordering may contribute to the properties of inhomogeneous states, such as quantized vortices.

Superfluidity and superconductivity are observed in liquids whose macroscopic behaviour is governed by quantum statistics; the latter becomes important at low temperatures when the thermal de Broglie wavelength of the particles in the medium exceeds the interatomic spacing. Superfluids can flow without friction through narrow gaps and tubes. Only two liquids, ^4He

and ^3He , are known to exhibit this property. Superconductivity, the loss of electrical resistance, is observed in many metals below a critical temperature T_c that varies from one substance to another. In superconductors, superfluidity is manifested by the conduction electrons.

Phenomenologically, a superfluid can be considered a liquid with two components: the normal fraction has density ρ_n and the superfluid fraction has density ρ_s . The superfluid behaves like a single macroscopic particle with a quantum-mechanical wavefunction $\Psi(\mathbf{r})$, usually called the order parameter. Physical properties of particular interest in superfluids are those in which the quantum character of the liquid manifests itself on a macroscopic scale. The most interesting of these are phenomena that appear in superfluid ^4He and ^3He under rotation. Due to their quantum nature, unlike classical liquids, superfluids cannot rotate as a whole. Instead, circulation perforates them in quanta of individual vortex lines^{1,2} (Fig. 1).

In 1972, Osheroff, Richardson, and Lee found the superfluid A and B phases of ^3He below 3 mK; the P/T (pressure/temperature) diagram is illustrated in Fig. 2. This discovery opened a new and exciting era in ultralow-temperature physics. In 1981, vortices in rotating superfluid ^3He were first investigated experimentally at the Helsinki University of Technology³.

During the past five years, intensive studies with the nuclear magnetic resonance (NMR) technique, persistent-current measurements, ion-mobility investigations, and hydrodynamic experiments have all revealed several new, unprecedented and often unexpected phenomena associated with vorticity in anisotropic superfluids. Space permits only a brief discussion of persistent-current experiments and a more detailed description of the pioneering NMR results; the latter have provided most of the fundamental information on vortices in superfluid ^3He . For example, in $^3\text{He-A}$, continuous distribution of vorticity has been observed, as in a classical liquid—accompanied, however, by a liquid-crystal-like texture. In $^3\text{He-B}$, new superfluid phases nucleate inside the cores of the quantized vortex lines—as revealed by the spontaneous vortex-core magnetization and the first-order vortex-core phase transition³ (see Fig. 2).

Vortices in classical superfluids

In ^4He , which has a nuclear spin $S=0$ and which obeys Bose-Einstein statistics, superfluidity is understood as the Bose condensation of individual atoms into the zero-momentum state^{4,5}. The Bose condensate may be described by a macroscopic wavefunction $\Psi(\mathbf{r}) = C(\mathbf{r})e^{i\Phi(\mathbf{r})}$, whose amplitude squared, $C^2(\mathbf{r})$, is the density, $\rho_s(\mathbf{r})$, of the superfluid component. The

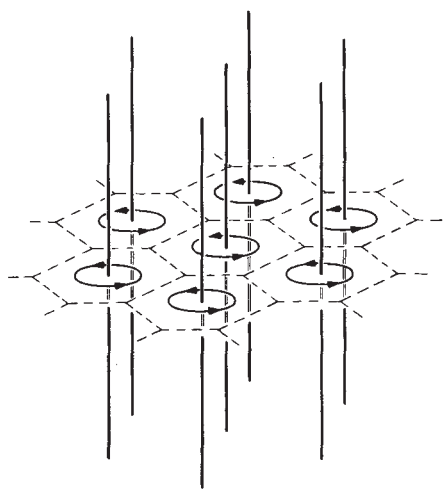


Fig. 1 The hexagonal lattice of quantized vortices, as seen by an observer co-rotating with the superfluid. The rotation axis is parallel to the array of vertical lines depicting vortex cores at the centre of each primitive lattice cell; lines with arrows illustrate streamlines of superflow. Dashed lines indicate the hexagonal vortex-lattice cell boundaries, along which the superfluid is at rest in relation to the rotating container. On average, the lattice of quantized vortices mimics solid-body rotation with a parabolic meniscus. Rotating superfluid ^4He has been studied since the 1940s¹, first by Andronikashvili² extensive measurements have been conducted recently by Glaberson, by Packard, by Reppy, by Rudnick, and by Tsakadze and their co-workers.

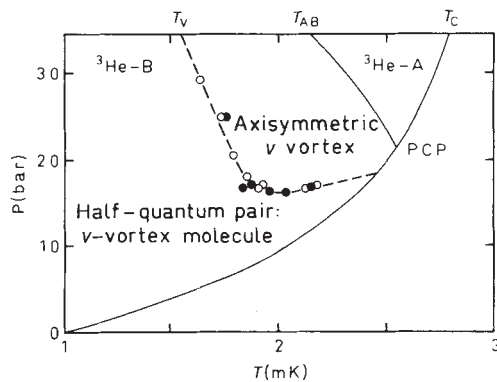


Fig. 2 Phase diagram of liquid ^3He . On rotating $^3\text{He-B}$ in an open-volume NMR cell, a transition is observed³ on crossing the dashed line T_v in the P, T -plane¹⁵. The intersection of the T_{AB} and T_c curves is the polycritical point, PCP. $\circ$, NMR experiments at $H = 284$ G; $\bullet$, $H = 568$ G. This vortex-core transition is a topological change between two inequivalent ways to accommodate vorticity^{48,49}; see Fig. 8.

gradient of the phase, $\nabla\Phi(\mathbf{r})$, gives the superflow velocity:

$$\mathbf{v}_s = \frac{\hbar}{m_4} \nabla\Phi(\mathbf{r}) \quad (1)$$

Being the gradient of a scalar, superflow is potential, curl-free: $\nabla \times \mathbf{v}_s = 0$. Experiments show that this is true everywhere, except along special vortex lines at which $\nabla\Phi$ is singular; the flow velocity diverges as $v_s \propto 1/r$ where r is the distance from the vortex core. The integral $\oint \mathbf{v}_s \cdot d\mathbf{r}$ along any closed path around a vortex is proportional to the change in phase $\Delta\Phi$, which—to keep the wavefunction single-valued—must equal an integral multiple of 2π . Hence the circulation of superflow is quantized in He II:

$$\oint \mathbf{v}_s \cdot d\mathbf{r} = m \frac{h}{m_4} \quad (2)$$

in quanta of $h/m_4 = 0.997 \times 10^{-3} \text{ cm}^2 \text{ s}^{-1}$; m_4 is the mass of a ^4He atom and m is an integer. Equation (2) resembles the Bohr-Sommerfeld quantization condition: the angular momentum per particle is quantized in units of $\hbar$.

Viscous fluids in a container rotating at the angular velocity Ω behave like a solid body, that is $\mathbf{v} = \Omega \times \mathbf{r}$ and $\nabla \times \mathbf{v} = 2\Omega$. This is not possible for a superfluid, but because rotational motion is energetically more favourable than the stationary nonrotating situation, a vortex lattice (Fig. 1), is created.

Each vortex normally carries only one quantum of circulation, because the energy of two vortex lines is smaller than the energy of a single one with two circulation quanta. Repelling each other, vortices form a hexagonal two-dimensional lattice. On average, such an arrangement imitates solid-body rotation. Consequently, the total number of singly quantized vortex lines in a vessel of cross-sectional area A is

$$N = \frac{2\Omega A m_4}{h} \quad (3)$$

and this corresponds to $N/A \approx 2,000$ vortices cm^{-2} at $\Omega = 1 \text{ rad s}^{-1}$.

Vortex lines in superfluid ^4He have no particular internal structure: the superfluid density tends to zero on the vortex axis, where the fluid is normal. Hence the kinetic-energy density of the rotating superfluid component, $\frac{1}{2}\rho_s v_s^2$, remains finite everywhere, in spite of the $1/r$ -singularity of $\mathbf{v}_s$ on the vortex axis. The characteristic scale for spatial changes in the order parameter, determining the radius of the vortex core, is the so-called coherence length, $\xi = 0.1\text{--}0.2 \text{ nm}$ in ^4He ; this is of the same order as the interatomic distance.

Superfluid phases of ^3He

Like electrons, ^3He atoms having a half-integer (nuclear) spin $S = \frac{1}{2}$ are governed by Fermi-Dirac statistics. Strong mutual interactions renormalize the ^3He atoms into quasiparticles but do not change the nature of the quantum statistics. According to the BCS theory of Bardeen, Cooper and Schrieffer, Fermi systems undergo a transition into a macroscopically coherent new superfluid ground state in the presence of an arbitrarily weak attraction between the quasiparticles. The long-range order may be viewed as being due to the formation of mutually overlapping giant molecules, called Cooper pairs.

In superconductors⁶, as a consequence of the isotropy of interactions between the conduction electrons, these pairs have angular momentum $L = 0$ (s-wave pairing) and total spin $S = 0$ (singlet). In superfluid ^3He , the Cooper-pair states⁷ have $L = 1$ (p-wave pairing) and $S = 1$ (triplet). The order parameter $\Psi(\mathbf{r})$ for a p-wave paired superfluid can be written as a linear combination of spherical harmonics $\chi_{S\mu}$ and $Y_{L\nu}$, respectively, for the states with different projections $S_z = \pm 1, 0$ ($\chi_{1\mu}$ with $\mu = \pm 1, 0$) and $L_z = \pm 1, 0$ ($Y_{1\nu}$ with $\nu = \pm 1, 0$) on directions of the quantization axes, $\hat{\xi}$ and $\hat{z}$:

$$\Psi(\mathbf{r}) = \sum_{\mu\nu} a_{\mu\nu} \chi_{1\mu} Y_{1\nu} \quad (4)$$

In $^3\text{He-A}$, all pairs are in the state with $L_z = 1$ and $S_z = 0$, that is, there is only one component, $a_{0+} \neq 0$. In $^3\text{He-B}$, with $L + S = J = 0$, states with $L_z = \pm 1, 0$ and $S_z = \pm 1, 0$ are equally populated, and there are three equal components of the order parameter: $a_{+-} = a_{00} = a_{-+}$.

In both the A and B phases^{8,9}, the average value of the spin, namely the spontaneous Cooper-pair magnetization, is zero. In $^3\text{He-B}$, the average of the angular momentum is also zero; hence this liquid is completely isotropic in the absence of inhomogeneities. In contrast, all Cooper pairs in $^3\text{He-A}$ share the same direction of $\mathbf{L}$, which determines the orbital anisotropy axis $\mathbf{l}(\parallel \hat{z})$. Therefore, many properties of $^3\text{He-A}$, for example viscosity, are different in directions parallel or perpendicular to $\mathbf{l}$. Hence $^3\text{He-A}$ is an anisotropic superfluid liquid crystal. In addition to the orbital anisotropy, there is a magnetic anisotropy in the A phase, described by the vector $\mathbf{d}$ which defines the spin quantization axis $\hat{\xi}$; the Cooper-pair spin $\mathbf{S}$ is perpendicular to $\mathbf{d}$. Because of this anisotropy, the paramagnetic susceptibility of the A phase is a uniaxial tensor.

Superfluidity in $^3\text{He-A}$ is not necessarily associated with curl-free flow because $\nabla \times \mathbf{v}_s$ may be different from zero. In fact, the A phase can support nonsingular vorticity in the presence of an inhomogeneous distribution of the anisotropy axis $\mathbf{l}$. Previously, curl-free flow was considered an essential property inherent to every superfluid. Thus it has become necessary to revise many theoretical concepts of quantum liquids, especially their behaviour under rotation.

Rotating cryostat

All NMR measurements—and most of the other types of experiments—on superfluid ^3He in the rotating equilibrium state have been made in the ROTA cryostat at Helsinki^{10,11}. This apparatus incorporates a copper nuclear stage, a dilution refrigerator, an activated charcoal cryoadsorption pump, radial and axial air bearings, and computerized measurement electronics. The apparatus can be rotated at angular velocities Ω up to 3.0 rad s^{-1} . The copper nuclear stage, weighing 4.0 kg, is first precooled to 20 mK, with the 8-T magnet on, by means of a dilution refrigerator. The nuclear stage is then thermally isolated by operating a superconducting heat switch and demagnetized in 10 h to 150 mT, whereby it cools to 0.4 mK. Next, the charcoal adsorption pump is switched on and the external pumping lines are disconnected, allowing the cryostat to be rotated.

The ^3He sample cell is on top of the nuclear stage and thermally connected to it. For NMR measurements the cell was a cylinder, 5 mm in diameter and 30 mm long. Its geometry is

thus clearly defined, and the NMR response of stationary ^3He in this configuration is well known. The measurements were made at three fixed frequencies; 1,845, 920 and 460 kHz. The magnetic field was usually oriented along the rotation axis but could be tilted sideways when required. The field was swept linearly in 2–3 min over the resonance at 56.9, 28.4 or 14.2 mT. For persistent-flow measurements, the experimental cell was a 50-mm-diameter torus, 6-mm thick; the flow path was filled with 20- μm plastic powder to enhance v_c , the critical velocity for superflow.

Persistent currents in ^3He

One of the characteristic properties of superfluids is their persistent-flow capability: in He II the flow of the superfluid component around a closed path persists as long as the liquid is kept under its transition temperature, $T_\lambda = 2.17$ K. Experiments for observing persistent currents in the superfluid phases of ^3He have been carried out by groups at Cornell University^{12,13} and in Helsinki^{14,15}. In both laboratories, an a.c. gyroscopic technique was used. Figure 3 illustrates the gyroscope employed in the Cornell experiments. The principle of the measurement is: when the ring is driven about the y -axis at the resonant frequency of the ω -mode, a sinusoidal response around the x -axis results, provided that there is an angular momentum in the ring which is caused by superfluid flow.

In the gyroscopic experiments, the sample was first cooled well below the superfluid transition temperature at rest. The cryostat was then rotated around the z axis at an angular velocity Ω and stopped, and the vibrational amplitude in the ω -direction was recorded. An angular momentum was detected in the torus only if the speed of rotation had exceeded a critical value Ω_c . In such conditions, an irreversible creation of superflow occurs which persists even after the cryostat has been brought to rest. In $^3\text{He-B}$, no decay of superflow was observed in Helsinki¹⁴ during 48 h, implying an effective viscosity at least 12 orders of magnitude less than that in the normal Fermi liquid at the same temperature: thus $^3\text{He-B}$ shares properties with 'classical' superfluids. In the Helsinki experiments no observable current was found to persist in $^3\text{He-A}$. However, the Cornell group¹³, exploiting a method with a higher quality factor near T_c , also detected persistent currents in $^3\text{He-A}$. A discontinuous change in the critical flow velocity, attributed to a vortex transition in $^3\text{He-B}$ was seen in the Helsinki experiments¹⁵.

Vortices in $^3\text{He-A}$

Before any experiments had been conducted on rotating $^3\text{He-A}$, it had been recognized on theoretical grounds that this phase should possess curious superflow properties^{16–18}. In particular,

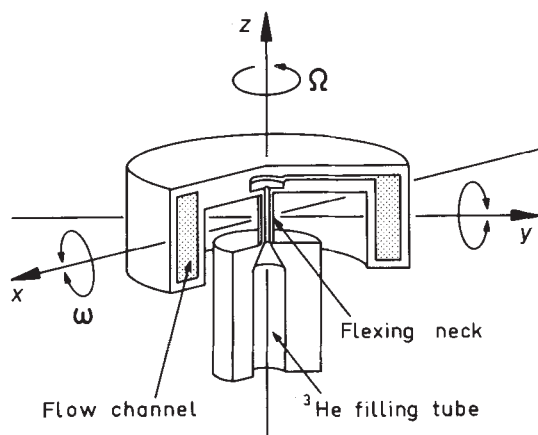


Fig. 3 Schematic diagram of the a.c. gyroscope employed in the Cornell experiments¹². To immobilize the normal fluid, the toroidal flow channel is filled with fibrous material; the resulting average open volumes have linear dimensions of the order of 0.1 mm. For operating details, see text.

the possible existence of continuous vortices—which were subsequently found in Helsinki—had first been predicted by Chechetkin¹⁹ and also discussed by Anderson and Toulouse²⁰. In addition, half-quantum vortices had been suggested by Volovik and Mineev²¹ and subsequently by Cross and Brinkman²² on the basis of the Cooper-pair structure in $^3\text{He-A}$.

Superfluidity of $^3\text{He-A}$

The p-wave pairing state of $^3\text{He-A}$ was discussed above: each Cooper pair in this liquid is under rotational motion with the projection $L_z = 1$ of the angular momentum on the direction of the quantization axis $\mathbf{l}$. This internal motion is mixed with the macroscopic Cooper-pair centre-of-mass motion. Because of this, the nature of superflow in $^3\text{He-A}$ is very peculiar: rotational motion can follow from the presence of spatially inhomogeneous distributions of the $\mathbf{l}$ vector¹⁷. These textures unbalance the internal Cooper-pair motion, producing a nonpotential net superflow without the formation of singularities.

The spatial structure of the order parameter is determined by competition among several forces. The elastic stiffness of the $\mathbf{l}$ and $\mathbf{d}$ fields tends to make them uniform; this gives ^3He its liquid-crystal-like behaviour. The dipolar interaction tries to lock $\mathbf{l}$ and $\mathbf{d}$ parallel or antiparallel to each other. At cell boundaries, $\mathbf{l}$ is perpendicular to the walls¹⁶. In the presence of superfluid flow, $\mathbf{l}$ tends to orient parallel to $\mathbf{v}_s$. An external magnetic field tries to lock $\mathbf{d}$ into the plane perpendicular to $\mathbf{H}$ ($\mathbf{d} \perp \mathbf{S} \parallel \mathbf{H}$).

The actual texture of the liquid depends on the interplay between these effects, with the configuration of lowest free energy prevailing. As the textures of $\mathbf{l}$ can create rotational superflow in $^3\text{He-A}$, the circulation of superflow velocity in the A phase need not be quantized. At first sight, the possibility of nonpotential superflow seems to admit purely solid-body rotational motion. However, the energetics of the $\mathbf{l}$ texture require that even in $^3\text{He-A}$ under rotation, a periodicity in the order-parameter texture appears.

Vortex structures

Every elementary vortex-lattice cell contains a vortex line with m quanta of circulation. The value of the circulation quantum in superfluid ^3He , $h/2m_3 = 0.662 \times 10^{-3} \text{ cm}^2 \text{ s}^{-1}$, differs from that in superfluid ^4He [see equation (2)] not only due to the different isotopic mass, but also owing to the factor 2, caused by Cooper pairing.

The possibility of rotational superflow means that a singular vortex line can be transformed into another one with continuous coreless distribution of vorticity. However, due to topological constraints²³, the elimination of the singularity in $\mathbf{v}_s$ on vortex lines with odd number of circulation quanta is accompanied by the appearance of a singularity in the $\mathbf{l}$ texture ($\nabla \mathbf{l} \rightarrow \infty$ on the axis). However, vortices with an even m can be transformed into vortices with a continuous distribution of vorticity, while maintaining a nonsingular distribution of the $\mathbf{l}$ texture. Hence, vortices with odd m are always singular, but the singularity in vortices with even m can be continuously dissolved.

The singular singly quantized vortex has two different cores, the hard core with radius of the order of the coherence length, $\xi \approx 10 \text{ nm}$, inside which the order parameter deviates from its bulk $^3\text{He-A}$ value, and the soft core, with a radius of the order of $\xi_D \approx 6 \mu\text{m}$, the dipolar healing length outside which the dipolar force is greater than the elastic forces. Beyond the soft core, $\mathbf{l}$ is practically uniform; all vorticity in the liquid is thus localized inside the soft core. The continuous vortices have no singular core; however, depending on the external conditions, they may acquire a soft core with localized vorticity. The coherence length in ^3He is about two orders of magnitude longer than that in ^4He .

A very different type of singular vortex is possible if one takes into account that the total wavefunction for Cooper pairs in

$^3\text{He-A}$ is the pure product of the orbital and spin parts: $\Psi(\mathbf{r}) = a_0 \chi_{10} Y_{1+}$ [see equation (4)], where the dependence on the coordinate $\mathbf{r}$ of the centre of Cooper-pair mass arises from the $\mathbf{r}$ -dependencies of the quantization axes: $\mathbf{l}(\mathbf{r})$ and $\mathbf{d}(\mathbf{r})$. The quantization of circulation in 'classical' superfluids is simply a reflection of the fact that the phase of the order parameter, upon circling a closed contour in the superfluid, can only change by $\pm 2\pi m$ because the wavefunction must be single-valued, thus $\Psi(\mathbf{r}) \rightarrow \Psi(\mathbf{r}) e^{\pm i 2\pi m} = \Psi(\mathbf{r})$. But in $^3\text{He-A}$, we may have for the orbital part of $\Psi(\mathbf{r})$: $Y_{1+}(\mathbf{r}) \rightarrow Y_{1+}(\mathbf{r}) e^{\pm i\pi}$, while for the spin part: $\chi_{10}(\mathbf{r}) \rightarrow \chi_{10}(\mathbf{r}) e^{\pm i\pi} [\mathbf{d}(\mathbf{r}) \rightarrow -\mathbf{d}(\mathbf{r})]$, resulting in: $\Psi(\mathbf{r}) \rightarrow \Psi(\mathbf{r}) e^{\pm i 2\pi} = \Psi(\mathbf{r})$. Hence, half-quantum vortices are possible. However, a half-quantum singularity in the orbital part of $\Psi(\mathbf{r})$ is not separable from a simultaneous singularity in the spin part. Thus in $^3\text{He-A}$ we may have an example of topological confinement of singularities that is closely related to similar phenomena in Grand Unified Theories.

Therefore, three topologically different types of vortices²³ can exist in superfluid $^3\text{He-A}$: (1) singular vortices with an odd quantum of circulation, and with topological charge 1; (2) continuous coreless vortices with even integer quanta of circulation, and with topological charge 0; and (3) singular vortices with a half-integer quantum of circulation, and topological charges $\frac{1}{2}$ and $-\frac{1}{2}$. The topological summation laws are: $\frac{1}{2} + \frac{1}{2} = 1$, $1 + 1 = 0$, and $\frac{1}{2} + 1 = -\frac{1}{2}$. This means, in particular, that two singular singly quantized vortices can combine to produce a doubly quantized continuous vortex—with consequent annihilation of the singularity.

In $^3\text{He-A}$ under rotation, a vortex lattice and a corresponding periodic $\mathbf{l}$ -vector texture²⁴ is present: there is a great difference in the quantization conditions for $^3\text{He-A}$ and for superfluids with traditional properties. In the latter case, circulation along any closed contour around the vortex is an integer in terms of the circulation quantum. For the vorticity in $^3\text{He-A}$, this is valid only along the boundary of a primitive cell. From energy considerations alone, at angular velocities of the order of $1\text{--}3 \text{ rad s}^{-1}$, it follows that, in the so far experimentally relevant magnetic

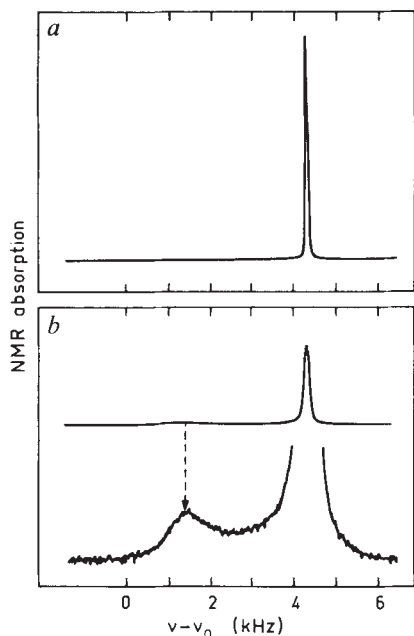


Fig. 4 NMR spectra¹¹ of stationary and rotating $^3\text{He-A}$. Additional broadening of the main peak, due to rotation, is apparent; the shift of this peak towards higher frequencies by 4.6 kHz from ν_0 is related to the spin-orbit coupling. Arrow points to the satellite absorption peak which appears in the NMR signal during rotation. It is caused by the excitation of spin waves trapped by the continuous vortex texture. Lowest curve is vertically magnified $\times 50$. $a, \Omega = 0$; $b, \Omega = 1.21 \text{ rad s}^{-1}$.

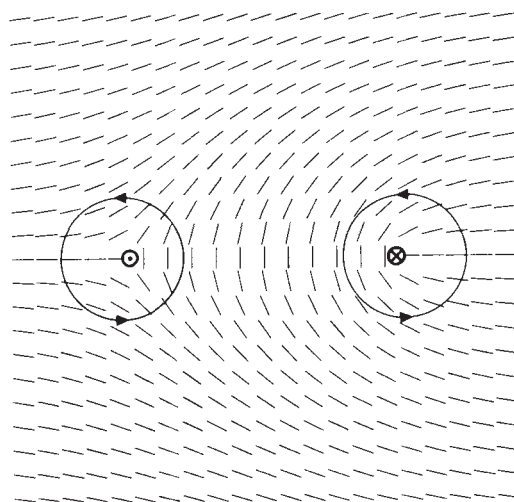


Fig. 5 Texture of the magnetic anisotropy vector $\mathbf{d}$ around a pair of vortices with half-quantum circulation²⁸ in $^3\text{He-A}$. Energy of the local disturbance ('soliton') in the uniform $\mathbf{d}$ texture between the half-quantum pair is proportional to their separation; the pair is thus glued together by the soliton string, as in quark confinement; see Fig. 8b. The liquid-crystal-like $\mathbf{d}$ texture displays defects with opposite Frank indices at the half-quantum vortices: $\mathbf{d}$ rotates by $+\pi$ and by $-\pi$ on circling the two singular points ($\odot$ and $\otimes$).

fields of 15–60 mT, the most advantageous structure in the rotating bulk $^3\text{He-A}$ is a lattice of singular singly quantized vortices.

NMR on vortices

How does one detect and investigate the physical properties of individual vortex lines in ^3He ? The most versatile tool used so far is the NMR technique, based on the resonance absorption of electromagnetic radiation at the Larmor frequency $\nu_L = \gamma H_0$, corresponding to the precession of nuclear magnetization in the liquid under a constant field H_0 ; γ is the gyromagnetic ratio of ^3He . Note that NMR cannot be used on ^4He because the nuclear spin $S = 0$ for this isotope.

Precession of the nuclear spins $\mathbf{S}$ in an anisotropic superfluid is not free: the directions of the $\mathbf{l}$ and $\mathbf{d}$ vectors tend to be locked by the spin-orbit coupling. This produces an additional torque on the spins of the Cooper pairs, and because of this a shift in the NMR absorption line away from γH_0 occurs, see Fig. 4.

In a strong magnetic field, the doubly quantized vortex texture is modified: If $\mathbf{H}$ is applied parallel to the axis of the rotating vessel, the $\mathbf{d}$ vector is in the plane perpendicular to $\mathbf{H} \parallel \Omega$ because of magnetic anisotropy. Due to the spin-orbit coupling, $\mathbf{l}$ tends to be parallel to $\mathbf{d}$, thus $\mathbf{l} \perp \mathbf{H}$. This results in the broken axisymmetry of the $^3\text{He-A}$ vortices in a magnetic field. The distribution of the $\mathbf{l}$ vector deviates from the homogeneous situation $\mathbf{l} \parallel \mathbf{d}$ only in the central part of every elementary vortex-lattice cell inside the soft core, where the energy of the inhomogeneity is larger than that due to the spin-orbit coupling. Thus the soft cores, which contain the vorticity, are well separated: at the typical rotational speed of 1 rad s^{-1} , the spacing between the vortices is $\sim 0.3 \text{ mm}$ and thus more than an order of magnitude larger than the diameter of the soft core.

The regions of inhomogeneous distribution of $\mathbf{l}$ act as potential wells for localized standing spin waves. Therefore, a periodic $\mathbf{l}$ -vector texture leads to the appearance of an additional satellite NMR peak in the absorption spectrum at the frequency of the local vortex mode (Fig. 4). The intensity of this peak is proportional to the number density of vortices, that is to the angular velocity of rotation [see equation (3)]. The calculated intensity of the vortex peak for a lattice of continuous vortices^{25–27}, each with two circulation quanta, is ~ 10 times smaller than that for a singular vortex lattice and is in fair agreement with the experi-

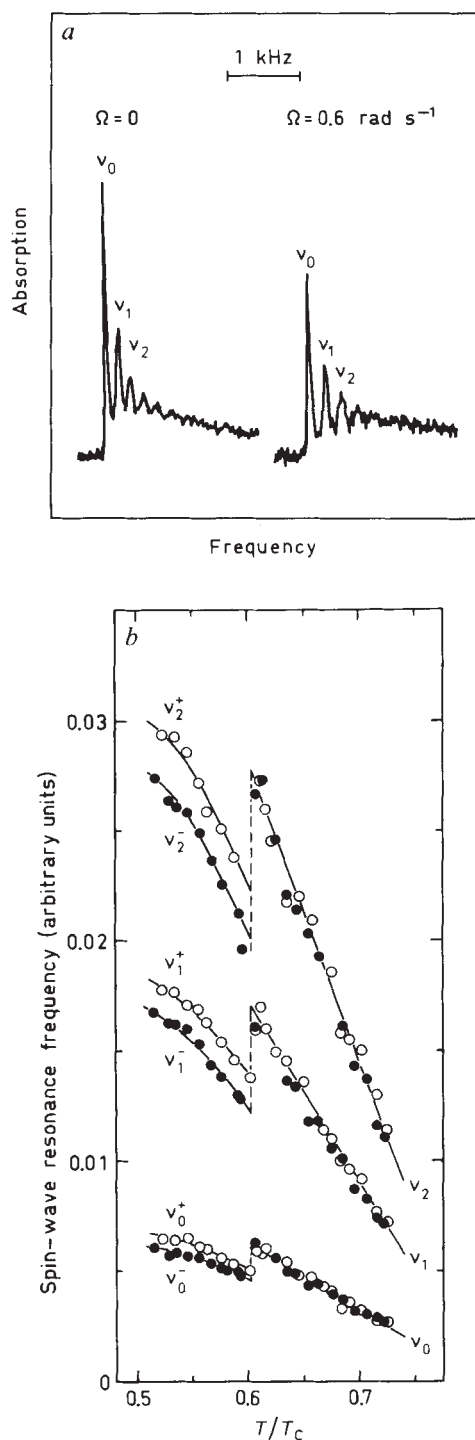


Fig. 6 *a*, Measured NMR absorption spectra¹¹, at 29.3-bar pressure, for $^3\text{He-B}$ as functions of frequency in the stationary case and at $\Omega = 0.6 \text{ rad s}^{-1}$. The envelope of the observed oscillatory curve corresponds to the 'flare-out' texture^{31,32}. The series of almost evenly spaced spin-wave satellites results from the coherent nuclear motion of the Cooper pairs. Separation between the resonant peaks at $\nu_0, \nu_1, \nu_2, \dots$ is larger during rotation than in the stationary liquid. *b*, The spin-wave resonance frequencies ν_0, ν_1 , and ν_2 , of Fig. 6*a* at $\Omega = 0.6 \text{ rad s}^{-1}$, as functions of reduced temperature. The data illustrate a gyromagnetic effect³⁵ due to the magnetic moment of the vortex cores^{36,37}: the resonance frequency measured during rotation with Ω parallel to the magnetic field $\mathbf{H}$ (○) differs from that for Ω antiparallel to $\mathbf{H}$ (●). The effect is large at $T < 0.6T_c$ but hardly observed for $T > 0.6T_c$; the vortices below and above this transition temperature at $T_c = 0.6T_c$ have different core structures and, therefore, different intrinsic magnetic moments.

mentally observed NMR satellite intensity¹¹. This supports the existence of continuous vortices in $^3\text{He-A}$, instead of the energetically slightly more advantageous singularly quantized vortices. A possible explanation for this is that in the creation of a singular vortex a high potential barrier has to be overcome. Therefore, we may conjecture that continuous vortices are formed first under rotation and that once they exist in metastable states it becomes difficult to transform one continuous vortex into two singular ones for creating a singular vortex array.

Half-quantum vortices have not been seen experimentally in $^3\text{He-A}$ (see, however, below and Figs 2 and 8*b*). These singularities may be observed in the A phase only when the spin and orbital variables can be considered independent. Such a situation exists in narrow gaps between parallel plates, where the energy of the inhomogeneity is larger than the spin-orbit coupling^{22,28}. Because of the peculiar topology, the doubly ($\pm$) valued $\mathbf{d}$ -vector field changes sign ($\mp$) on circling a half-quantum vortex line, see Fig. 5. This produces an interesting consequence²⁹: the equation for the spin wavefunction bound by the inhomogeneous $\mathbf{d}$ -vector texture is analogous to the Schrödinger equation for a charged particle in a 'magnetic field', concentrated in the vortex-core tube. This analogue of the Aharonov-Bohm effect also generates spin wave scattering, which should result in significant additional broadening of the main NMR line, possibly providing an experimental signature allowing identification of half-quantum vortices.

Vortices in $^3\text{He-B}$

In 1978, when the decision was made to build a rotating ^3He cryostat in Helsinki, it was believed that the vortex lines in superfluid $^3\text{He-B}$ would be just like those in superfluid ^4He , with no internal structure. In the first NMR experiments³, however, a spontaneous phase transition in the core was discovered. This immediately implied a complicated vortex core; it is appreciated that vortices in superfluid ($^3\text{He-A}$ and) $^3\text{He-B}$ provide a beautiful application for the concept of broken symmetry³⁰ in condensed-matter physics. The vortex-core transition has been established to involve a change in the topology of the vortex-core matter between two distinct ways that vorticity can flare out from real to momentum space.

NMR spectroscopy

In $^3\text{He-B}$, the mutual orientation of the quantization axes of $\mathbf{L}$ and $\mathbf{S}$ is arbitrary in the absence of spin-orbit coupling; the superfluid is thus degenerate with respect to rotations of the spin axis $\hat{\mathbf{z}}$ about the orbital axis $\hat{\mathbf{z}}$. Any three-dimensional rotation may be represented by an angle θ around some axis specified by a unit vector $\hat{\mathbf{n}}$. The B-phase states are degenerate with respect to directions of $\hat{\mathbf{n}}$ and, in the absence of spin-orbit coupling, to the value of θ . Moreover, as in superfluid ^4He , the $^3\text{He-B}$ states are also degenerate with respect to the value of the overall phase Φ of the order parameter. This means that the energy of the homogeneous liquid does not depend on $\hat{\mathbf{n}}$, θ or Φ . In inhomogeneous $^3\text{He-B}$, for example, when vortices are present, the energy depends on the gradients of $\hat{\mathbf{n}}$, θ and Φ .

As in ^4He , the superflow velocity in $^3\text{He-B}$ is determined by the gradient of the phase: $\mathbf{v}_s = (\hbar/2m_3)\nabla\Phi$ and the circulation of superflow is quantized. Therefore, there are no coreless vortices in $^3\text{He-B}$ and the superfluidity of this phase is in several respects analogous with the properties of superfluid ^4He . In particular, a two-dimensional lattice of singly quantized vortices is formed under rotation.

In $^3\text{He-B}$, the order parameter and the texture are fully determined by the $\hat{\mathbf{n}}$ vector, which is analogous to the directrix vector describing the orientation of molecules in liquid crystals. Because the orientational force exerted by an external magnetic field on the nearly isotropic $^3\text{He-B}$ is so small, the magnetic healing length ξ_H is long; effects due to walls are thus very important in $^3\text{He-B}$ and they give rise to $\hat{\mathbf{n}}$ textures^{31,32}. Inside the vortex cores, however, the magnetic anisotropy energy is of

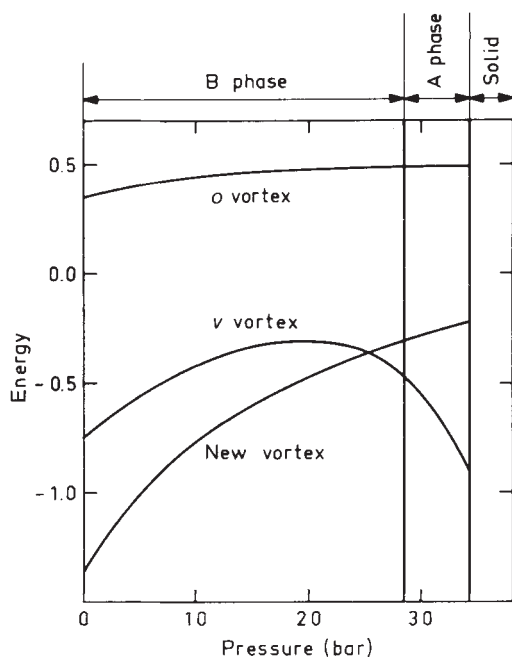


Fig. 7 Energies of the axisymmetric o and v vortices, and the nonaxisymmetric new vortex, as functions of pressure according to numerical calculations⁴¹. Pressures are given on a theoretical scale: for example, the tricritical point (see Fig. 2), actually at 21 bar, is at 28.5 bar.

the same order of magnitude as in $^3\text{He-A}$; therefore, vortices have a strong influence on the $\hat{n}$ texture. The v_z field around the vortex also has an orientating effect on $\mathbf{n}$ (ref. 33).

The most powerful experimental tool for observing vortices in $^3\text{He-B}$ is the NMR technique, as in $^3\text{He-A}$. The precessing motion of the nuclear spins is coupled with the order-parameter distribution, determined by the spatial variations of $\hat{n}$, θ and Φ . The energy minimum for spin-orbit coupling fixes θ to the 'magic' angle $\theta_0 = 104^\circ$. Another consequence of the coupling is the appearance of a small anisotropy, caused by the magnetic energy term which orients $\hat{n}$ along $\mathbf{H}$. In the presence of an axial magnetic field, as a result of boundary conditions on $\hat{n}$ and the magnetic anisotropy, the equilibrium orientation only occurs in the bulk liquid at distances $> \xi_H[\text{cm}] \approx 4(1 - T/T_c)/H[\text{mT}]$ from the walls.

In a cylindrical container, with an axially oriented $\mathbf{H}$, a spiral-like 'flare-out' texture results^{31,32}. As in $^3\text{He-A}$, the $\hat{n}$ -texture gives rise to observable localized standing spin waves. These stationary coherent nuclear spin modes are confined by the flare-out texture (Fig. 6a). The frequencies of these modes depend on the spatial variation of $\hat{n}$. With the use of NMR it is thus possible to register changes in $\hat{n}$. Consequently, if the texture of the liquid is altered by rotation, this can be observed. The influence of rotation on $\hat{n}$ is due to those vortex-core regions where the balance between the equal-weight projections of $S_z = \pm 1, 0$ and $L_z = \pm 1, 0$ is broken. Because of this, the cores of vortices in $^3\text{He-B}$ possess an anisotropic magnetic susceptibility³³, which depends on the orientation of $\hat{n}$ and on the direction of Ω .

The appearance of an additional energy of magnetic anisotropy due to the vortex cores suggests the possibility of studying the vortices by the measurement of the Ω dependence of the frequencies of the spin-wave modes in the flare-out texture, see Fig. 6. The signature of vortices is also in the shift of the NMR signal³⁴ in a field $\mathbf{H}$ tilted with respect to the direction of the axis of rotation: the shift is zero without the vortices. In addition, a gyromagnetic effect³⁵ has been observed, which was manifested by the fact that the cores of the $^3\text{He-B}$ vortices possess a magnetic moment³⁶⁻³⁹; for further discussion, see Fig. 6b and below.

Vortex structures

In superfluid ^3He , the hard vortex-core size is in the Ginzburg-Landau (GL) regime near T_c of the order of several coherence lengths, $\xi_{GL} \approx 10(1 - T/T_c)^{-1/2}$ nm, which is the temperature-dependent Cooper-pair extent. Because of the large size of vortices and the complicated nature of ordering it is not necessary to have normal liquid inside the vortex cores in $^3\text{He-B}$; instead, other combinations of different p-wave pairing states may nucleate. The structure of the core can be determined by numerical minimization of the vortex energy.

In the cylindrical coordinate frame³⁶, the order-parameter amplitudes $a_{\mu\nu}(\mathbf{r})$ in equation (4) may be written for a singly quantized vortex as

$$a_{\mu\nu}(\mathbf{r}) = \begin{pmatrix} C_{++} e^{-i\phi} & C_{+0} & C_{+-} e^{i\phi} \\ C_{0+} & C_{00} e^{i\phi} & C_{0-} e^{2i\phi} \\ C_{-+} e^{i\phi} & C_{-0} e^{2i\phi} & C_{--} e^{3i\phi} \end{pmatrix} \quad (5)$$

For an axisymmetric vortex, the above amplitudes $C_{\mu\nu}$ are only functions of the radial distance r from the vortex axis and independent of the azimuthal angle ϕ . The core structures have been classified using symmetry arguments^{37,38}: it was found that an axisymmetric vortex in $^3\text{He-B}$, with unit quantum of circulation, may be in five different states, distinguished by the internal discrete symmetry of the vortex core. These states are labelled o , u , v , w and uvw ; in all of them the vortex core possesses a magnetic moment. The most symmetric o vortex, first discussed by Ohmi *et al.*³⁶, has a normal core with superfluidity destroyed on the axis, as in ^4He . In the u , v and w vortices the internal symmetry is partially destroyed, whereas in the uvw vortex it is completely broken.

Energetically the most favourable of the different possible axisymmetric vortices is the v vortex with broken parity^{37,38}; in this case, all the nine functions $C_{\mu\nu}$ in equation (5) are real. Although the B-phase pairing amplitudes C_{+-} , C_{00} , and C_{-+} also in this vortex must tend to zero on the vortex axis, because of broken parity, this vortex, nevertheless, has a superfluid core: the two components without phase factors in equation (5), the A phase C_{0+} and a ferromagnetic so-called β phase C_{+0} , are not prohibited by any centrifugal barrier at the vortex axis. This results in the core which practically consists of the A phase, as it is energetically more favourable than the β phase. These results were later independently confirmed by other workers³⁹⁻⁴¹.

Because of the ferromagnetic superfluid in the core, a large spontaneous vortex magnetization $\mathbf{M}$ results; its direction is determined by the directions of Ω and $\hat{n}$. This vortex magnetization, created by rotation, gives rise to a new free-energy term, $-\mathbf{H} \cdot \mathbf{M}$, which is odd in both Ω and $\mathbf{H}$ and which explains the gyromagnetic effect shown in Fig. 6b. A change in the sense of rotation or in the direction of the magnetic field changes the texture of the liquid. Therefore, $\mathbf{M}$ could be measured³⁵ directly by NMR spectroscopy even though its value in $^3\text{He-B}$ rotating at $\Omega = 1 \text{ rad s}^{-1}$ is tiny: only of the order of 10^{-11} nuclear Bohr magnetons per ^3He atom.

The v vortex possesses a spontaneous electric polarization, also caused by the violation of parity in the core³⁸. The w vortex would also show striking properties: it has a complex superfluid core and it displays spontaneous superflow along the vortex axis. The uvw vortex shares the properties of both the v and w vortices.

Vortex-core transition in $^3\text{He-B}$

Perhaps the most interesting feature signalling a nontrivial vortex-core structure in $^3\text{He-B}$ is the phase change, which takes place when crossing a transition line while the liquid is rotating, see Fig. 2. This was observed experimentally as abrupt discontinuities in the spacing of standing spin waves³ and as a jump in the vortex-core magnetization³⁵ (Fig. 6b) and as a jump in the critical flow velocity^{14,15}. These changes are independent of the angular velocity of rotation, that is; the density of vortices,

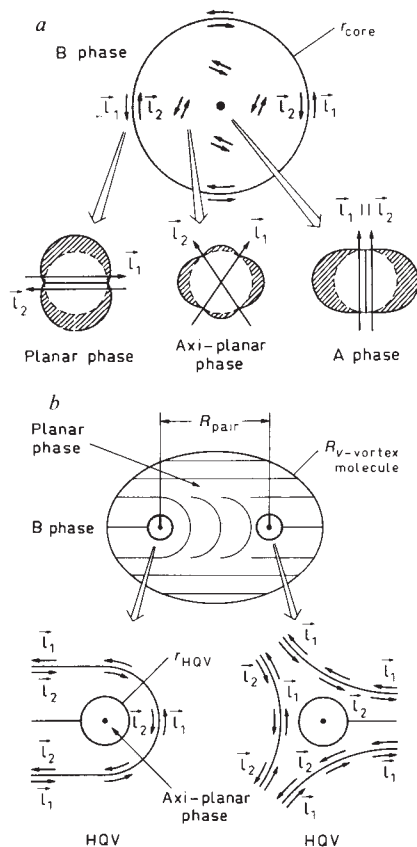


Fig. 8 Topology of the energy gap (hatched areas) on the Fermi sphere in the cross-sectional plane of v vortices in $^3\text{He-B}$ at high and low pressures (see Fig. 2). *a*, In the axisymmetric v vortex³⁸, point vortices first appear in the 'planar-like' phase at $r = r_{\text{core}}$. For $r < r_{\text{core}}$ vorticity is supported by continuous winding, in the 'axi-planar-like' phase, of the nodes for quasiparticles with spin up (l_1) and down (l_2), changing from antiparallel at $r = r_{\text{core}}$ to parallel at $r = 0$ with 'A-phase-like' core. *b*, The nonaxisymmetric v -vortex 'molecule'. The two centres are both a half-quantum vortex⁴⁹ (HQV) confined in the core (see Fig. 5). Each HQV is accompanied by a disclination in the $\mathbf{l}$ -vector field with half-integer winding number, reversing l_1 into l_2 on circling either HQV. The nodes lie in the cross-sectional plane ('planar-like' phase), except inside inner cores of radius r_{HQV} where they flare out of the plane and deviate from each other (in the 'axi-planar-like' phase). This 'bag' model of the core 'molecule' resembles hadrons with the HQVs corresponding to quarks.

implying that the transition is not affected by interactions between vortices. Rather, the transition is attributed to a change in the core of isolated vortex lines.

Much work has been devoted to identifying the vortices associated with this unexpected transition. First, it was contemplated^{37,42} that one might understand the transition to be one between two of the five axisymmetric vortices. It was found, however, that the v vortex provides the absolute energy minimum among the axisymmetric vortices³⁸. Also, it is stable against small nonaxial perturbations⁴³. Salomaa and Volovik⁴⁴ suggested that the mechanism of the vortex-core transition in $^3\text{He-B}$ is breaking of axisymmetry. Nonaxisymmetric vortex structures were also considered by Theodorakis and Fetter⁴⁵.

Recently, it has been reported^{41,46,47} that a nonaxisymmetric vortex has a lower energy than the axisymmetric v vortex at low pressures. Thuneberg⁴¹ was the first to perform the full two-dimensional energy minimization by solving numerically the Ginzburg-Landau differential equations; his results are illustrated in Fig. 7, whereas Volovik and Salomaa^{46,47} only carried out investigations within sectors of definite symmetry. At all pressures the energy of the axisymmetric v vortex is clearly

higher than the energies of the axisymmetric v vortex³⁷ or the nonaxisymmetric 'double core' new vortex with two points of local maxima in the energy density. At low pressures, the new vortex has the lowest energy. With an increase of pressure, the energy difference between the new vortex and the axisymmetric v vortex decreases and their energies are equal at ~ 3 bar below the tricritical pressure; this is in excellent agreement with the experimental data of Fig. 2. Thuneberg's calculations show that the axisymmetric v vortex becomes stable at higher pressures.

Volovik and Salomaa^{46,47}, on the basis of their systematic study³⁸ of all the discrete broken symmetries, confirmed Thuneberg's⁴¹ calculation and identified the new vortex as the nonaxisymmetric v vortex of their classification scheme. At present it seems that in $^3\text{He-B}$ the nonaxisymmetric v vortex is stable at low pressures and the axisymmetric v vortex at high pressures (Figs 2 and 7). Volovik and Salomaa find that the mechanism of restructuring the superfluid core in the vortex transition is independent of any model-dependent details: the first-order vortex-core transition is triggered by the proximity of the A phase alone, and driven by the broken parity of the v vortex, combined with superfluid condensation-energy considerations.

Point vortices on the Fermi surface

What is the actual physical meaning of the vortex-core transition: what happens to superfluid vorticity? To answer these questions, it is crucial to appreciate that vortices in ^3He can couple the real $\mathbf{r}$ and the momentum $\mathbf{p}$ spaces: superfluid-core vortices can escape the real-space singularity on the vortex axis by flaring out into the momentum space; this process is associated with nodes in the energy gap on the Fermi surface within the vortex core⁴⁸. We denote by l_1 and l_2 the directions of the nodes for the two energy gaps, $\Delta_1(\hat{\mathbf{p}})$ and $\Delta_2(\hat{\mathbf{p}})$, for the two possible ^3He spin projections. In both v vortices, there appear two pairs of such nodes. These objects are topologically interesting point vortices ('boojums') in the $\mathbf{p}$ -space, around which the phase Φ_i of $\Delta_i = |\Delta_i(\hat{\mathbf{p}})| e^{i\Phi_i(\hat{\mathbf{p}})}$ (for $i = 1, 2$) changes by $\pm 2\pi$; therefore, the singularity ($v_s \propto 1/r$) can be resolved without breaking superfluidity on the vortex axis.

Volovik and Mineev⁴⁸ were the first to propose that the core transition in $^3\text{He-B}$ involves a change in the topology of the vortex-core matter; however, the transition proved more interesting⁴⁹ than that originally suggested. To escape the singularity in the axisymmetric v vortex³⁸ (Fig. 8a), l_1 and l_2 each cover half the Fermi sphere: unwinding unit quantum of circulation thus results according to: $m = \frac{1}{2} + \frac{1}{2} = 1$; this is associated with the simultaneous spatial variation of each orbital anisotropy axis, l_1 and l_2 , just like vorticity can be supported by the continuous winding of the order parameter in the $\mathbf{l}$ textures of $^3\text{He-A}$.

For the nonaxisymmetric v vortex, the topological scenario for flaring-out of the singularity is essentially different⁴⁹. The composite core 'molecule' is illustrated in Fig. 8b. It contains two half-quantum vortices, 'quarks'⁴⁷, coupled by 'strings': disclinations in both l_1 and l_2 , closely resembling the $\mathbf{d}$ -field disclination in Fig. 5. One may here verify that each point vortex covers one quarter of the Fermi surface in both of the half-quantum vortices: $m = (\frac{1}{4} + \frac{1}{4}) + (\frac{1}{4} + \frac{1}{4})$; hence the two 'quarks' do represent objects with fractional topological charge. Consequently, Fig. 2 identifies the vortices involved in the core transition, whose first-order character signals the topological bifurcation in the distribution of vorticity into the momentum space.

Conclusions

The varied vortex-core structures in the superfluid phases of ^3He lead to the possibility of several phase transitions inside the cores. These can occur in the hard vortex core, but also in the much more extended soft-core region. In $^3\text{He-B}$, so far only one of many possible transitions has been found in the NMR spectra^{3,15,35}; the nature of this vortex-core transition is understood theoretically⁴⁹. The search and study of these new physical

phenomena will be the subject of future theoretical and experimental investigations.

To date, these studies, principally with use of the NMR technique, have concentrated on the static properties of vortices. In the future, problems related to the dynamics of vortices need to be tackled: nucleation, the mechanisms for transitions—possibly involving monopoles—to lower-energy states, and the formation and properties of vortex rings. In this research, ions⁵⁰ have already been used and experiments involving ultrasound are being planned.

The physics of vortex phenomena in ^3He is interesting from a much broader point of view, owing to the different mechanisms for topological confinement, exemplified by the generation of the d-vector soliton (string) that glues together the pairs of half-quantum vortex disclinations in $^3\text{He-A}$, as in quarks. There are many conceptual connections to quantum field theories, and close analogies with defects in liquid crystals⁵¹. The structure of the vortex core in the B phase is intimately related to the topology of the point vortices on the Fermi surface⁴⁸. The novel dimerization mechanism in flaring-out of vorticity for the non-axisymmetric v vortex molecule suggests the identity of the vortex-core transition in $^3\text{He-B}$: the nonaxisymmetric v vortices are topologically confined pairs of half-quantum vortices, analogous to quarks inside nucleons. Provided that experiments will confirm this identification⁴⁹, vortices with half-integer circulation have first been found in rotating superfluid $^3\text{He-B}$.

It is commonly believed that the neutron mass in pulsars undergoes a BCS-type transition to form a superfluid Fermi liquid in a $^3\text{P}_2$ state⁵². Normal-core vortices with spontaneous magnetization have been discussed for neutron stars^{53,54}, and the analogue experiments on the superfluidity of pulsars⁵⁵ have been conducted on rotating superfluid ^4He . Being a Fermi liquid, also paired in a p-wave state, superfluid ^3He should provide a better substance for such model experiments. It is likely that, just as in superfluid ^3He , vortices in pulsars also display a superfluid core. This possibility could introduce a substantial revision in present understanding of heat transport in neutron stars and thereby in our picture of the latter stages of supernovae.

An interesting experiment was recently suggested by Zurek⁵⁶,

the proposal is to carry out a rapid isothermal expansion of liquid ^4He through the superfluid transition, whereby vorticity should be generated at the interfaces of the superfluid condensation fronts. Such vortex lines are proposed to be analogous to the cosmic strings^{57,58} that are believed to have nucleated in the phase transitions of the early Universe. An analogue model experiment on superfluid ^3He (here presumably by a rapid isothermal compression through the superfluid transition) would be more difficult to carry out. However, this is potentially more interesting because the 'vacuum' states of superfluid ^3He , with the most complicated known broken symmetry in condensed matter, provide a more relevant analogue system for Grand Unified Theories. Here the vortex density can also be directly measured by the NMR technique. This idea has also been proposed for liquid crystals (M. Yu. Khlopov and S. Obukhov, unpublished data), where the creation of disclination lines at the phase transition may be seen directly with a microscope.

Vortices in heavy-fermion superconductors remain largely unexplored. Provided that these materials⁵⁹ do exemplify p-wave or other unconventional pairing, it is possible that some of the conceptual generalizations which have been necessary to make progress in understanding experimental results on vortices in superfluid ^3He will also prove useful in the context of heavy-fermion superconductors.

In conclusion, vortices in superfluid ^3He , although highly interesting for their own sake also seem to be important in a broader context beyond low-temperature physics; for more detailed expositions of the topic, see refs 34, 43 and 60 (review).

We thank P. Berglund, Yu. M. Bun'kov, A. L. Fetter, I. A. Fomin, A. D. Gongadze, P. J. Hakonen, O. T. Ikkala, S. T. Islander, G. A. Kharadze, M. Krusius, K. Maki, N. D. Mermin, T. Ohmi, R. E. Packard, J. P. Pekola, J. D. Reppy, J. A. Sauls, N. Schopohl, J. T. Simola, the late J. S. Tsakadze and especially G. E. Volovik for collaboration at various stages of the work reported here, and for helpful comments. This research has resulted from the ROTA project, supported by the USSR Academy of Sciences, by the Commission for Scientific and Technical Collaboration between Finland and the USSR, and by the Academy of Finland.

1. Tilley, D. R. & Tilley, J. *Superfluidity and Superconductivity* (Adam Hilger, Bristol, 1985).
2. Andronikashvili, E. L. & Mamaladze, Yu. G. *Rev. Mod. Phys.* **38**, 567–625 (1966).
3. Ikkala, O. T. *et al.* *Pis'ma Zh. eksp. teor. Fiz.* **35**, 338–341 (1982); *JETP Lett.* **35**, 416–420 (1982).
4. Feynman, R. P. in *Progress in Low Temperature Physics* Vol. I, Chap. 2 (ed. Gorter, C. J.) (North-Holland, Amsterdam, 1964).
5. Putterman, S. J. *Superfluid Hydrodynamics* (North-Holland, Amsterdam, 1974).
6. Tinkham, M. *Introduction to Superconductivity* (McGraw-Hill, New York, 1975).
7. Leggett, A. J. *Rev. Mod. Phys.* **47**, 331–414 (1975).
8. Wheatley, J. C. *Rev. Mod. Phys.* **47**, 415–470 (1975).
9. Lee, D. M. & Richardson, R. C. in *The Physics of Liquid and Solid Helium* Vol. II, Chap. 4 (eds Bennemann, K. H. & Ketterson, J. B.) (Wiley, New York, 1978).
10. Hakonen, P. J. *et al.* *Cryogenics* **23**, 243–250 (1983).
11. Hakonen, P. J., Ikkala, O. T., Islander, S. T., Lounasmaa, O. V. & Volovik, G. E. *J. Low Temp. Phys.* **53**, 425–476 (1983).
12. Gammel, P. L., Hall, H. E. & Reppy, J. D. *Phys. Rev. Lett.* **52**, 121–124 (1984).
13. Gammel, P. L., Ho, T.-L. & Reppy, J. D. *Phys. Rev. Lett.* **55**, 2708–2711 (1985).
14. Pekola, J. P., Simola, J. T., Nummilla, K. K., Lounasmaa, O. V. & Packard, R. E. *Phys. Rev. Lett.* **53**, 70–73 (1984).
15. Pekola, J. P. *et al.* *Phys. Rev. Lett.* **53**, 584–587 (1984).
16. Ambegaokar, V., deGennes, P. G. & Rainer, D. *Phys. Rev. A* **9**, 2676–2685 (1974); erratum *Phys. Rev. A* **12**, 345 (1975).
17. Mermin, N. D. & Ho, T.-L. *Phys. Rev. Lett.* **36**, 594–597 (1976); erratum *Phys. Rev. Lett.* **36**, 832 (1976).
18. Volovik, G. E. *Usp. Fiz. Nauk* **143**, 73–109 (1984); *Soviet Phys. Usp.* **27**, 363–384 (1984).
19. Chechekin, V. R. *Zh. eksp. teor. Fiz.* **71**, 1463–1474 (1976); *Soviet Phys.—JETP* **44**, 766–772 (1976).
20. Anderson, P. W. & Toulouse, G. *Phys. Rev. Lett.* **38**, 508–511 (1977).
21. Volovik, G. E. & Mineev, V. P. *Pis'ma Zh. eksp. teor. Fiz.* **24**, 605–608 (1976); *JETP Lett.* **24**, 561–563 (1976).
22. Cross, M. C. & Brinkman, W. F. *J. Low Temp. Phys.* **27**, 683–686 (1977).
23. Volovik, G. E. & Mineev, V. P. *Zh. eksp. teor. Fiz.* **72**, 2256–2274 (1977); *Soviet Phys.—JETP* **45**, 1186–1196 (1977).
24. Volovik, G. E. & Kopnin, N. B. *Pis'ma Zh. eksp. teor. Fiz.* **25**, 26–28 (1977); *JETP Lett.* **25**, 22–24 (1977).
25. Seppälä, H. K. & Volovik, G. E. *J. Low Temp. Phys.* **51**, 279–290 (1983).
26. Seppälä, H. K. *et al.* *Phys. Rev. Lett.* **52**, 1802–1805 (1984).
27. Zotos, X. & Maki, K. *Phys. Rev. B* **30**, 145–151 (1984).
28. Volovik, G. E. & Salomaa, M. M. *Zh. eksp. teor. Fiz.* **88**, 1656–1663 (1985); *Soviet Phys.—JETP* **61**, 986–990 (1985).
29. Salomaa, M. M. & Volovik, G. E. *Phys. Rev. Lett.* **56**, 1313 (1986).
30. Anderson, P. W. *Basic Notions of Condensed Matter Physics* (Benjamin, Menlo Park, California, 1984).
31. Smith, H., Brinkman, W. F. & Engelsberg, S. *Phys. Rev. B* **15**, 199–213 (1977).
32. Maki, K. & Nakahara, M. *Phys. Rev. B* **27**, 4181–4185 (1983).
33. Gongadze, A. D., Gurgenshvili, G. E. & Kharadze, G. A. *Fiz. Nizk. Temp.* **7**, 821–825 (1981); *Soviet J. Low Temp. Phys.* **7**, 397–398 (1981).
34. Bun'kov, Yu. M., Gurgenshvili, G. E., Krusius, M. & Kharadze, G. E. *Usp. Fiz. Nauk* **144**, 141–168 (1984); *Soviet Phys. Usp.* **27**, 731–747 (1984).
35. Hakonen, P. J. *et al.* *Phys. Rev. Lett.* **51**, 1362–1365 (1983).
36. Ohmi, T., Tsuneto, T. & Fujita, T. *Prog. theor. Phys.* **70**, 647–653 (1983).
37. Salomaa, M. M. & Volovik, G. E. *Phys. Rev. Lett.* **51**, 2040–2043 (1983).
38. Salomaa, M. M. & Volovik, G. E. *Phys. Rev. B* **31**, 203–227 (1985).
39. Passvogel, T., Tewordt, L. & Schopohl, N. *J. Low Temp. Phys.* **56**, 383–414 (1984).
40. Hasegawa, Y. *Prog. theor. Phys.* **73**, 1258–1260 (1985).
41. Thuneberg, E. V. *Phys. Rev. Lett.* **56**, 359–362 (1986).
42. Fetter, A. L. *Phys. Rev. B* **31**, 7012–7017 (1985).
43. Fetter, A. L. in *Progress in Low Temperature Physics* (ed. Brewer, D. F.) Vol. X (North-Holland, Amsterdam, in the press).
44. Salomaa, M. M. & Volovik, G. E. *Am. Inst. Phys. Proc. No. 103*, 210–228 (1983).
45. Theodorakis, S. & Fetter, A. L. *J. Low Temp. Phys.* **52**, 559–591 (1983).
46. Volovik, G. E. & Salomaa, M. M. *Pis'ma Zh. eksp. teor. Fiz.* **42**, 421–424 (1985); *JETP Lett.* **42**, 521–524 (1985).
47. Salomaa, M. M. & Volovik, G. E. *Phys. Rev. Lett.* **56**, 363–366 (1986).
48. Volovik, G. E. & Mineev, V. P. *Zh. eksp. teor. Fiz.* **83**, 1025–1037 (1982); *Soviet Phys.—JETP* **56**, 579–586 (1982).
49. Salomaa, M. M. & Volovik, G. E. *Europhys. Lett.* **2**, 781–788 (1986).
50. Simola, J. T. *et al.* *Phys. Rev. Lett.* **57**, 1923–1926 (1986).
51. Kléman, M. *Points, Lines and Walls in Liquid Crystals, Magnetic Systems and Various Ordered Media* (Wiley, New York, 1983).
52. Pines, D. & Alpar, M. A. *Nature* **316**, 27–32 (1985).
53. Muzikar, P., Sauls, J. A. & Serene, J. W. *Phys. Rev. D* **21**, 1494–1502 (1980).
54. Sauls, J. A., Stein, D. L. & Serene, J. W. *Phys. Rev. D* **25**, 967–975 (1982).
55. Tsakadze, J. S. & Tsakadze, S. J. *J. Low Temp. Phys.* **39**, 649–688 (1980).
56. Zurek, W. H. *Nature* **317**, 505–508 (1985).
57. Zel'dovich, Ya. B., Kobzarev, I. Yu. & Okun, L. B. *Zh. eksp. teor. Fiz.* **67**, 3–11 (1974); *Soviet Phys.—JETP* **40**, 1–5 (1975).
58. Vilenkin, A. *Phys. Rep.* **121**, 263–315 (1985).
59. Fisk, Z., Ott, H. R., Rice, T. M. & Smith, J. L. *Nature* **320**, 124–129 (1986).
60. Salomaa, M. M. & Volovik, G. E. *Rev. Mod. Phys.* (submitted).

Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes

Claudio W. Pikielny*, Brian C. Rymond† & Michael Rosbash†

* Department of Biochemistry and † Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

Three splicing complexes formed with a yeast pre-messenger RNA during in vitro splicing can be resolved by non-denaturing gel electrophoresis after incubation in the presence of non-specific competitor RNA. The time course of the appearance of these complexes and their composition suggest that they represent an ordered pathway of splicing complex assembly.

THE detailed analysis of RNAs generated during pre-mRNA splicing *in vitro*¹⁻⁵ and *in vivo*⁶⁻⁸ has led to the definition of a two-step splicing pathway in mammalian systems as well as in the yeast *Saccharomyces cerevisiae* (for recent reviews see refs 9, 10). In the first step, the 5' splice junction is cleaved, generating two molecules, the exon 1, and the intron-exon 2, in a lariat configuration, due to a 2'-5' linkage between the 5' end of the intron and the 2'-OH of an adenosine in the intron (which in *Saccharomyces cerevisiae* is the last A of the TACTAAC box). In the second step the two exons are ligated and the intron is excised as a lariat. The two-step pathway appears similar for *S. cerevisiae* and mammalian systems but differences in the *cis*-acting signals and the effects of mutations in these *cis*-acting sequences, indicate that some aspects of pre-mRNA splicing are different¹¹⁻¹³.

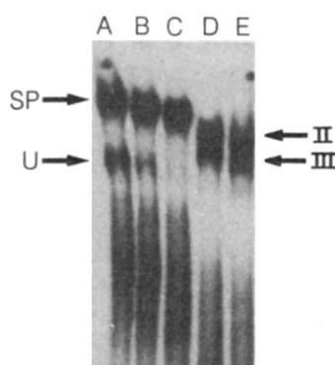


Fig. 1 Effect of non-specific competitor RNA on the electrophoretic resolution of splicing complexes. A splicing reaction was carried out as described^{4,5,11} with recombinant plasmids transcribed *in vitro* with SP6 polymerase²⁹ and containing rp51A sequences¹³. The reaction was incubated for 20 min at 25 °C, stopped by the addition of Q buffer (400 mM KCl, 2 mM Mg(OAc)₂, 20 mM EDTA and 100 mM hydroxyethyl piperazine ethane sulphonic acid HEPES-NH₄OH (pH 7.5)), and separated into 10 µl aliquots which were analysed by electrophoresis on a non-denaturing agarose-acrylamide composite gel¹³. Just before loading, each 10 µl sample was incubated on ice for 10 min in the presence of a preparation of total yeast RNA³⁰ at a concentration of: lane A, 0 µg µl⁻¹; B, 0.1 µg µl⁻¹; C, 0.2 µg µl⁻¹; D, 0.4 µg µl⁻¹; E, 1 µg µl⁻¹. The optimal concentration of non-specific RNA is ~1 µg µl⁻¹; higher concentrations did not improve the resolution but caused streaking and aggregation (not shown). Other nucleic acids have been tested as competitors, but neither transfer RNA nor DNA improve the resolution of splicing complexes. *Xenopus laevis* total RNA was just as effective as was *Saccharomyces cerevisiae* total RNA. The complexes observed, in the different lanes, labelled SP, U, II and III.

Splicing occurs in very large RNP (ribonucleoprotein) complexes¹⁴⁻¹⁶. In mammalian systems these splicing complexes, also called spliceosomes, contain snRNPs (small nuclear ribonucleoproteins^{15,17}) which are required for the splicing reaction^{9,10}. Point mutations in either of the two sequences essential for the splicing of yeast pre-mRNAs (the 5' splice junction and the TACTAAC box) inhibit spliceosome formation^{14,18}. It has only recently been possible to show that specific trimethyl-capped RNAs (for a review of small nuclear RNAs, see ref. 19) are present in yeast spliceosomes¹³ and that at least some of these are essential for the splicing reaction¹⁹.

We recently used non-denaturing polyacrylamide gels to analyse yeast RNPs formed during *in vitro* splicing¹³. Of the two RNPs visible, the first (called SP for spliceosome) shows all the characteristics of a splicing complex or spliceosome: its formation is dependent on (1) time, (2) ATP and (3) an intact TACTAAC box and 5' splice junction; it also contains the products of the first step of the splicing reaction (exon 1 and intron-exon 2 lariat intermediate) and a product of the second step, the excised intron in a lariat configuration. SP band formation precedes any detectable cleavage or ligation events, reaching a maximal level shortly after the addition of ATP. The second complex (named U for unspecific) behaves as a nonspecific RNP by all the above criteria. These data imply that only the SP complex is a spliceosome and that its formation is an early and probably essential step in the splicing of pre-mRNA.

Competitor RNA improves resolution

The addition of non-specific competitor RNA to a splicing reaction before electrophoresis increases the mobility of both the SP and the U complexes, presumably because loosely bound factors are lost (Fig. 1). The U band is readily dissociated with increasing concentrations of competitor RNA, as expected of a nonspecific complex. The SP band is less affected, implying that some interactions in this complex are resistant to the addition of competitor RNA. The SP band was resolved into two major bands of higher mobility, suggesting that in the absence of competitor RNA the SP band is heterogeneous and contains at least two different splicing complexes. The two bands visible in this experiment are bands II and III, the only ones normally detectable after a 20 min incubation (Fig. 1).

Three splicing complexes

A time course shows that three distinct complexes, called I, II and III in order of increasing electrophoretic mobility, are formed (Fig. 2, w.t.). In this typical experiment, substantial amounts of bands I and III were visible after 2 min of incubation; the concentrations were maximal at 5 min and decreased thereafter. Substantial amounts of bands I and III are formed after 30 s of incubation (data not shown), well before the formation of splicing intermediates and products can be detected (see

ref. 13, for example). Band II is not visible until 5 min of incubation after which it continues to increase in intensity. The first appearance of band II coincides with that of the intermediates of the reaction. The order of appearance of the three complexes suggests that two early complexes (bands I and III) may be intermediates in a pathway leading to the formation of the late complex (band II).

Complex formation with point mutant

Analysis of a strong point mutation at the TACTAAC box (the 3'-III mutation, TACTACC²¹) supports and extends this idea. As expected from *in vivo* studies²¹, incubation of an RNA substrate containing this mutation in an *in vitro* splicing reaction, results in no detectable intermediates or products (data not shown). Surprisingly, however, there was a substantial accumulation of band III (Fig. 2, 3'-III; note that this portion of the gel reads right to left), which formed more slowly than with a wild-type substrate RNA. No band I or band II was detected. The data suggest that the 3'-III mutant is blocked after formation of the early complex (band III) and hence that band III is the initial complex and leads to the subsequent formation of bands I and II with a wild-type substrate.

Distribution of labelled RNAs

To characterize these splicing complexes further, we analysed the RNA distribution across RNP gels (Fig. 3). Precursor RNA has a broad distribution with a clear maximum which coincides with band III and extends into band II and the region corresponding to band I (band I is indistinct at incubation times of 20 min). The products of the first step, the intron-exon 2 lariat and exon 1, are found almost exclusively in the band II fractions. The intron produced in the second step is distributed over bands II and III; in Fig. 3 it is more prominent in the band III region. A very low level of the exon 1-exon 2 RNA is detectable in band II.

We used a smaller RNA to determine the influence of substrate size on the migration of the different splicing complexes and to improve resolution. The rp51A $\delta 2$ RNA is only 247 nucleotides (nt) long (as opposed to 800 nt for rp51A) but is spliced accurately and efficiently *in vivo*²² and *in vitro*¹¹. The smaller substrate gives rise to a similar pattern of bands; the early forming complexes are designated bands III and I and the later complex is named band II (Fig. 4a). All three of these complexes show much higher mobility than the corresponding complexes formed with the larger substrate. There is also a fourth complex of higher mobility, not visible at early times (band IV). The RNAs present in each band from the 20 min time point were analysed on a denaturing polyacrylamide-urea gel (Fig. 4b); the distribution of radioactive RNA species between the bands is similar to that found for the larger transcripts (Fig. 4b and data not shown), confirming the correspondence of the complexes formed with the two transcripts. Band I contains mostly pre-mRNA and band III contains only pre-mRNA, whereas band II contains substantial amounts of both products of the first step, the intron-exon 2 lariat and exon 1. Bands I and II are not completely separated, but it can be seen that band I contains mostly pre-mRNA and that the products of the first step contribute >50% of the radioactivity in band II. Band IV contains mostly intron-exon 2 lariat, some excised intron and pre-mRNA, but no detectable exon 1 (compare lanes II and IV). It probably results from the partial dissociation of band II and cannot be an active complex because it lacks exon 1.

snRNAs associated with each complex

We have recently shown through partial purification of *in vitro* generated splicing complexes (or spliceosomes) by two different methods (non-denaturing electrophoresis and oligo(dT)-cellulose chromatography of spliceosomes formed on a polyadenylated substrate) that at least three different trimethyl cap-bearing snRNAs, 160, 185 and 215 nt long are associated with

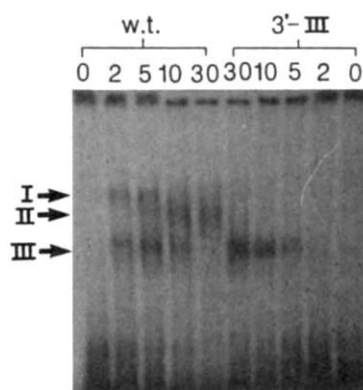


Fig. 2 Time course of formation of splicing complexes. Splicing reactions were carried out for the times indicated above each lane (in minutes), quenched and incubated in the presence of $1 \mu\text{g } \mu\text{l}^{-1}$ of total yeast RNA for 10 min on ice just before loading as in Fig. 1. (Note that, for the mutant transcript, the time points read from right to left.) w.t., splicing complexes formed with the wild-type rp51A transcript. 3'-III, splicing complexes formed with a transcript identical to the one in *a* except for an A to C transversion at the last A of the TACTAAC box²¹. The different bands observed are labelled I, II and III.

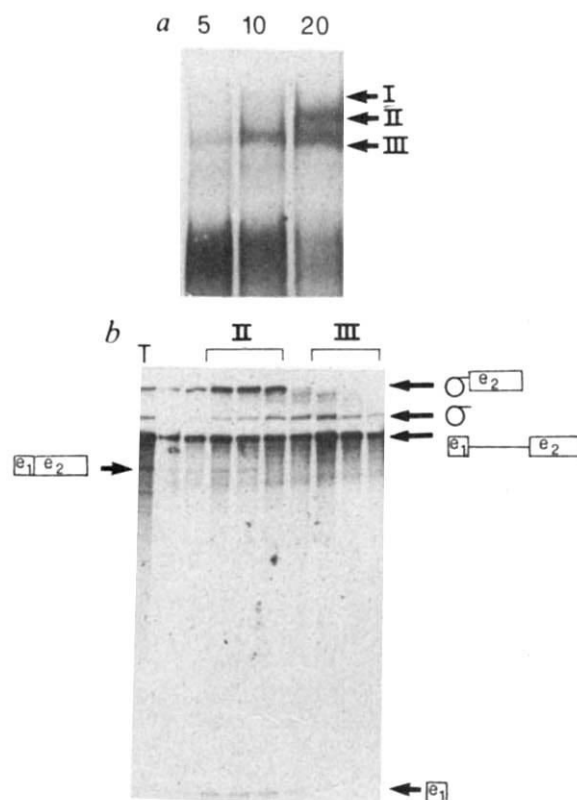


Fig. 3 Distribution of pre-mRNA, intermediates and products among splicing complexes formed with a long substrate. *a*, Splicing reactions with saturating amounts of substrate ($1 \text{ ng } \mu\text{l}^{-1}$) were incubated for 5, 10 and 20 min as indicated and resolved on a non-denaturing gel after incubation with total yeast RNA. The gel was then electroblotted to a DEAE membrane, which was then autoradiographed¹³. *b*, The lane corresponding to the 20-min time point was cut into 2-mm slices, the RNA eluted from each slice and analysed on a 5% polyacrylamide, 8 M urea gel¹³. The regions corresponding to bands II and III (as visible in *a*) are indicated; the lane between bands II and III which has the least amount of radioactivity and the two lanes corresponding to the two fractions above band II were not assigned to either band. T, unfractionated reaction. The structure of the different RNAs, as deduced from their electrophoretic mobility on this and other gels of different polyacrylamide concentrations, is shown at the side.

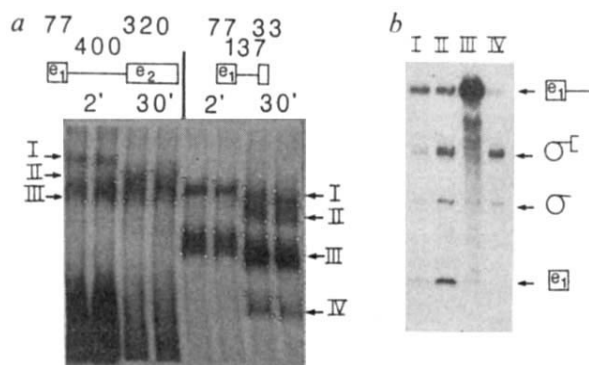


Fig. 4 Distribution of pre-mRNA, intermediates and products among splicing complexes formed with a short substrate. *a*, Preparative, non-denaturing gel showing the analysis of splicing complexes formed in the presence of a full-length (rp51A, 800 nt long) or short (rp51A $\delta 2/DdeI$, 247 nt long) RNA substrate after 2-min or 20-min incubations in a splicing reaction (each reaction was loaded on two lanes). An autoradiogram obtained after electroblotting the gel onto a DEAE membrane¹³ is shown. The structure of the transcripts is shown above the lanes. *b*, The RNAs were extracted from each band and analysed on a 5% polyacrylamide, 8 M urea gel¹³. The structure of the different RNAs as determined by their electrophoretic mobility is shown on the right.

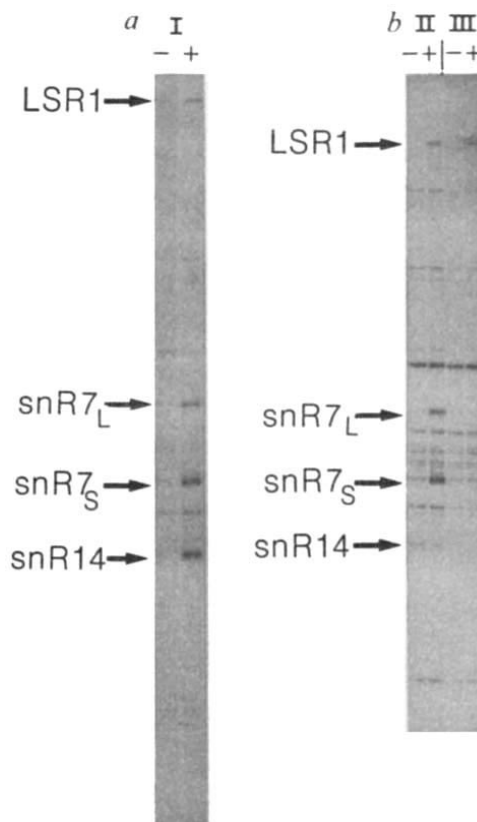


Fig. 5 Trimethyl-capped RNAs in the three splicing complexes; snRNAs present in bands I, II and III. Preparative reactions in the presence of the short substrate (rp51A $\delta 2/DdeI$) were incubated for 2 and 30 min at 25 °C and run on gels similar to those shown in Figs 3 and 4, after incubation for 10 min on ice in the presence of 1 $\mu\text{g } \mu\text{L}^{-1}$ of periodate-treated *Xenopus laevis* RNA (see below). The gels were then electroblotted onto DEAE membranes and the RNAs present in band I (panel *a*) or bands II and III (panel *b*) were eluted from each lane, ethanol-precipitated in the presence of 5 μg of salmon sperm DNA (see below), end-labelled with pCp and RNA ligase (as described³¹ with the modifications described in ref. 13), immunoprecipitated with anti-trimethyl-cap rabbit antibody (as described³², with the modifications described in ref. 13) and analysed on a 7% polyacrylamide 8 M urea gel. To ensure that the capped RNA species are specifically associated with *in vitro* assembled complexes, a reaction containing no exogenously-added substrate was processed in parallel¹³. The lanes containing complexes I, II and III are indicated: + and - denote the presence or absence of exogenously-added substrate in the splicing reaction. The sizes of the various RNAs were determined (on this and other gels) by comparison with 5S RNA, 5.8S RNA, the rp51A $\delta 2/DdeI$ transcript (247 nt), the rp51A transcript (800 nt) and 18S RNA. Two modifications were introduced in these experiments to ensure that the snRNAs observed were present in splicing complexes and not artefactual. First, *Xenopus laevis* RNA that had been oxidized with periodate (to make it unavailable for end-labelling with pCp and RNA ligase³¹) was used as competitor instead of total yeast RNA. The protocol for periodate treatment was provided by M. Green. Briefly, 200–500 μg of total *Xenopus laevis* RNA were resuspended in 1 ml of 0.3 M NaOAc pH 5.3, 1 ml of (4 mM NaIO₄, 10 mM NaOAc pH 5.3) was added and the mixture was incubated in the dark at 4 °C for 1 h, then at room temperature for 1 h. The reaction was quenched by adding 0.2 ml of propylene glycol (1,2-propanediol) and incubated at room temperature for an additional 10 min. The RNA was then phenol-extracted, ethanol-precipitated and washed extensively. Second, instead of tRNA as carrier in the ethanol precipitations after elution from the DEAE membrane, we used salmon sperm DNA which had been previously freed of any contaminating RNA by overnight incubation in 0.3M NaOH and ethanol precipitation in the presence of 2.5 M NH₄Ac, because we found that salmon sperm DNA was not (or only very poorly) labelled in a standard pCp end-labelling reaction.

yeast spliceosomes¹³. Comparison of our patterns of immunoprecipitated snRNAs with those in a recently reported study²³ suggests that these three species are those known as snR14 (160 nt), snR7_s (185 nt) and snR7_L (215 nt). The latter two are transcripts from the SNR7 gene and vary in the lengths of their 3' ends²³.

An additional, unusually large, trimethyl-capped species (1,100–1,200 nt) also appeared to be associated with complexes formed on a polyadenylated substrate, but not with spliceosomes purified by gel electrophoresis, probably because of a high background level in that region of the gel¹³. In the presence of carrier RNA the splicing complexes migrate in a region of the gel showing less background, allowing us to confirm that this RNA is also part of the spliceosome. Based on its molecular weight, it is almost certain that this large RNA corresponds to a recently discovered essential RNA species, lsr1 (large spliceosomal RNA1), the product of the LSR1 gene, which has extensive sequence homology to the mammalian U2 snRNA²⁴. Comparison of gel patterns suggests that this species is identical to snR20²³. Although these comparisons are not yet definitive, we will refer to these RNAs as LSR1, snR7_L, snR7_s and snR14.

We analysed the distribution of these RNAs among the different splicing complexes, comparing their level in each complex with the background level in the absence of *in vitro* assembled complexes (see Fig. 5 and ref. 13). By this criterion, each complex contains a specific complement of trimethyl-capped RNA species. lsr1 is clearly present in all three complexes; band III appears to contain only this species, as determined by this assay. Band I also contains the three other RNAs previously shown to be associated with spliceosomes, (snR7_s, snR7_L and snR14). Band II contains lsr1, snR7_s and snR7_L but is missing snR14. Thus, we can account for all the species described previously and the three complexes show different snRNA compositions (Fig. 6).

Pathway of complex formation

The addition of high concentrations of competitor RNA before electrophoresis greatly improves the resolution of splicing complexes over our previously reported method¹³ and allows a more detailed description than has been possible with the more conventional technique of sedimentation analysis^{14–17}. The competitor RNA presumably removes non-specific or relatively loosely-associated factors, reducing the size and increasing the electrophoretic mobility of the splicing complexes. The relative

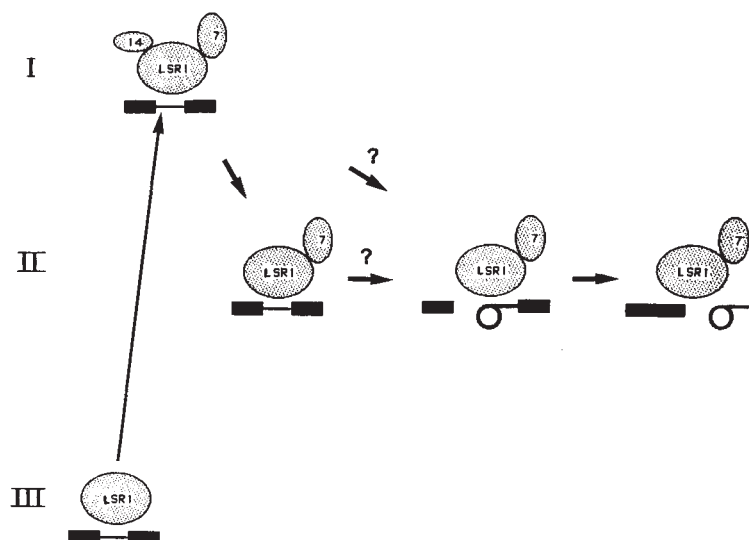


Fig. 6 Schematic summary of the data corresponding to the complexes formed on the pre-mRNA and those containing the products of the first and second step. Each band (I, II and III) is indicated with the complex(es) it contains. The composition in snRNAs is based on the assumption that a given band is homogeneous, which may not be the case. The arrows indicate transformation of one complex into another, which might involve other undetected intermediate complexes. The question marks reflect the fact that it is not clear which complex is the site of the first step.

mobility of the three major complexes is inversely related to the number of snRNAs found associated with each of them. It is not clear whether this is a causal relationship as the snRNA characterization includes a number of uncertainties. There may be additional snRNAs, undetected in this and our previous analysis (see discussion on this point in ref. 13) and differences in protein composition and/or conformation between the three complexes.

We suggest that the three complexes form in an ordered manner: III $\rightarrow$ I $\rightarrow$ II. To demonstrate such a pathway rigorously, one would have to show that it is possible to 'chase' pre-mRNA through the different complexes. This has not yet been possible but three lines of evidence, taken together, suggest that this interpretation is correct.

First, the kinetics of appearance of the various complexes show clearly that bands I and III appear early in the reaction, preceding any covalent modification of the substrate. Band II appears subsequently, as the products of the first step, 5' exon and lariat intermediate, first become detectable. Although the precise kinetics depend on the extract used, temperature and other details of the individual experiment, the two early bands (III and I) always appear together under normal conditions. In addition, both bands usually decrease substantially as band II intensifies (Fig. 2a). In several experiments in which the substrate pre-mRNA was saturating, band III accumulated to high levels and did not decrease (Figs 3a and 4a), even after long incubations (data not shown). In contrast, band I is only easily detectable at early times of incubation. Second, an examination of assembly under non-optimal conditions indicates that only band III is formed. This is true with a substrate containing a single point mutation at the branchpoint (Fig. 2b) as well as with defective extracts obtained, for example, by diluting a normal extract (data not shown). The III $\rightarrow$ I $\rightarrow$ II sequence explain the data most simply but a model where either III or I can lead to II is also possible. Analysis of the snRNAs associated with each complex suggests that III is the initial complex as it contains Lsr1 but not snR7 (S or L) or snR14. Band I is then formed with the binding of these two other snRNAs, and band II follows with the loss of snR14.

Sites of the two steps of splicing

The complex in which the first step occurs is expected to contain pre-mRNA and the products of this reaction. Both bands I and II are candidates for this active complex since we cannot resolve the two complexes clearly, they both appear to contain the two types of molecules. (The products of the first step are much more abundant in band II than band I, however.) If band I is the site of the first step, 5' cleavage and lariat formation are accompanied or rapidly followed by the loss of snR14 and the

transition to band II, which then more slowly undergoes the second step. If band II is the site of the first step, the presence of pre-mRNA in band II may be due to the fact that, at least *in vitro*, this active complex can also be formed (for example, band I can lose snR14) without undergoing the first step. We therefore cannot determine whether the loss of snR14 and the concomitant band I $\rightarrow$ band II transition occur before or after the first step of the reaction.

The site of the second step of the reaction, 3' cleavage and exon ligation, can be assigned more easily to band II. The products of the first step (intron-exon 2 lariat and exon 1), which are the substrates for the second step, are found in highest concentration in band II. The distribution of the products of the second step is also compatible with this view. In contrast to the lariat intron, almost none of the exon 1-exon 2 RNA is found associated with the large splicing complexes^{13,14,16,17}. In Fig. 3 a low level of exon 1-exon 2 RNA is detected and coincides with band II. A similarly low level of excised intron is also present in the band II fractions; larger amounts of this molecule are in faster migrating complexes which may be derived from band II simply by loss of the exon 1-exon 2 RNA.

This description does not relate the complexes observed to the *in vivo* complexes, or even to the complexes formed in *in vitro* splicing, from which bands I-III are derived. The treatment involved is particularly severe, including high salt, the addition of EDTA and, notably, incubation with non-specific competitor RNA (see Fig. 1) and can be expected to disrupt all but the strongest interactions. Thus the RNP gels show a 'core' splicing complex. This may affect the interpretation of the proposed assembly pathway. As an extreme possibility, there may exist a preformed spliceosome containing all the different snRNPs that will bind to the pre-mRNA; the interactions between the different snRNPs and/or between certain snRNPs and a given part of the pre-mRNA substrate might then become stronger or weaker as the splicing reaction proceeds. The different complexes observed here would then reflect different levels of interaction rather than the step by step association or disassociation of snRNPs with a complex containing the pre-mRNA substrate. Further studies may lead to an understanding of the interactions which produce stable complexes and the active splicing complexes, and the identification of the factors which recognize the conserved sequence elements in yeast introns.

This study was supported by grant GM23549 from the NIH to M.R. C.W.P. was supported by a Dretzin predoctoral fellowship and B.C.R. by a post-doctoral fellowship from the NIH. We thank A. Jacquier and M. Green for helpful discussions, D. Torrey for technical assistance, R. Lührmann for anti-TMG antibodies, and T. Tishman for typing the manuscript.

1. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. *Cell* **38**, 317-331 (1984).
2. Padgett, R. A., Konarska, M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. *Science* **225**, 898-903 (1984).
3. Konarska, M. M., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. *Nature* **313**, 552-557 (1985).
4. Lin, R.-J., Newman, A. J., Cheng, S.-C. & Abelson, J. *J. biol. Chem.* **260**, 14780-14792 (1985).
5. Newman, A. J., Lin, R.-J., Cheng, S.-C. & Abelson, J. *Cell* **42**, 335-344 (1985).
6. Domdey, H. *et al. Cell* **39**, 611-621 (1984).
7. Rodriguez, J. R., Pikielny, C. W. & Rosbash, M. *Cell* **39**, 603-610 (1984).
8. Zeitlin, S. & Efstratiadis, A. *Cell* **39**, 589-602 (1984).
9. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. *A. Rev. Biochem.* **55**, 1119-1150 (1986).
10. Green, M. R. *A. Rev. Genetics* **20**, 671-708 (1986).
11. Rymond, B. C. & Rosbash, M. *Nature* **317**, 735-737 (1985).
12. Ruskin, B., Pikielny, C. W., Rosbash, M. & Green, M. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2022-2026 (1986).
13. Pikielny, C. W. & Rosbash, M. *Cell* **45**, 869-877 (1986).
14. Brody, E. & Abelson, J. *Science* **228**, 963-967 (1985).
15. Grabowski, P. J., Seiler, S. R. & Sharp, P. A. *Cell* **42**, 345-353 (1985).
16. Frendewey, D. & Keller, W. *Cell* **42**, 355-367 (1985).
17. Bindereif, A. & Green, M. R. *Molec. cell. Biol.* **6**, 2582-2592 (1986).
18. Vijayraghavan, U. *et al. EMBO J.* **5**, 1683-1695 (1986).
19. Reddy, R. & Busch, H. in *Progress in Nucleic Acids Research and Molecular Biology* **30**, (Cohn, W. E., ed.) 127-162 (Academic, New York, 1983).
20. Cheng, S.-C. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2387-2391 (1986).
21. Jacquier, A. & Rosbash, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5835-5839 (1986).
22. Pikielny, C. W., Teem, J. L. & Rosbash, M. *Cell* **34**, 395-403 (1983).
23. Riedel, N., Wise, J. A., Swerdlow, H., Mak, A. & Guthrie, C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8097-8101 (1986).
24. Ares, M. Jr. *Cell* **47**, 49-59 (1986).
25. Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A. & Steitz, J. A. *Cell* **33**, 509-518 (1983).
26. Black, D. L., Chabot, B. & Steitz, J. A. *Cell* **42**, 737-750 (1985).
27. Chabot, B., Black, D. L., LeMaster, D. M. & Steitz, J. A. *Science* **230**, 1344-1349 (1985).
28. Ruskin, B. & Green, M. R. *Cell* **43**, 131-142 (1985).
29. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
30. Pikielny, C. W. & Rosbash, M. *Cell* **41**, 119-126 (1985).
31. England, T. E., Bruce, A. G. & Uhlenbeck, O. C. *Meth. Enzym.* **65**, 65-74 (1980).
32. Bringmann, P. *et al. J. biol. Chem.* **258**, 2745-2747 (1983).

LETTERS TO NATURE

An accretion event in the Seyfert galaxy NGC 5548

B. M. Peterson & G. J. Ferland

Department of Astronomy, The Ohio State University, 174 West 18th Avenue, Columbus, Ohio 43210, USA

Accretion on to a massive collapsed central object is now commonly thought to power active galactic nuclei (AGN)¹⁻³. If this is indeed so, then changes in luminosity of the system should correspond to changes in accretion rate, or, equivalently, the mass flux at regions a few Schwarzschild radii from the putative black hole. The observational consequences of such a change in accretion rate may not be pronounced, as gas near the central source of radiation is expected to be very highly ionized and difficult to observe outside the X-ray band. Here we draw attention to a remarkable change which occurred in the spectrum of the low-luminosity, optically variable, Seyfert galaxy NGC 5548. An increase in optical luminosity was accompanied by the appearance of a strong, abnormally broad He II 4,686-Å emission line. The appearance of this line indicates the presence of a component of highly ionized gas that did not contribute to the optical spectrum before the increase in luminosity. The greater line width suggests that the gas producing the He II emission is much closer to the central object than the broad region (BLR). This observation may be the first direct evidence of the accretion onto a central object.

We have been observing the low-luminosity Seyfert 1 galaxy NGC 5548 for some time⁴. Optical spectra have been obtained at regular intervals. A major goal of this work was to determine time lags between continuum activity and the response of the emission-line region. The time delay, which establishes the separation between the central engine and the BLR, is found to be around 31 days, corresponding to $r \approx 7.9 \times 10^{16}$ cm (ref. 5). If the BLR gas is bound in the potential well of the central object then the corresponding virial mass is $\sim 3 \times 10^7 M_\odot$.

During the course of these observations, the optical continuum increased by around 66% (see Fig. 1, in which the flux density of the continuum near 5,500 Å is shown, after removal of the underlying galaxy continuum). Details of the observational procedure and reduction techniques are discussed elsewhere⁴. This increase in visual brightness was accompanied by a similar increase in the extreme ultraviolet (EUV, that is $\lambda < 912$ Å) because the flux in the Balmer lines also increased (when the 31 day time-lag is taken into account). The fact that the increase in brightness in the continuum and Balmer lines increased together suggests that the luminosity of the EUV and optical continua are well correlated.

The surprising feature of these observations is the unexpectedly large increase in the strength of the He I 4,686-Å line (see

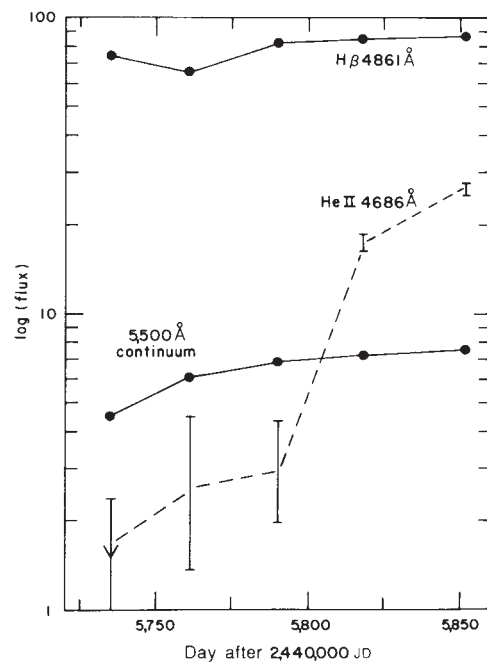


Fig. 1 The flux in H β , He II 4,686 Å and the 5,500 Å continuum, over the course of our observations. Line fluxes are in units of 10^{-14} erg s $^{-1}$ cm $^{-2}$ and continuum fluxes are in units of 10^{-15} erg s $^{-1}$ cm $^{-2}$ Å $^{-1}$.

Fig. 1). This phenomenon has not been reported before, although it is only within the last few years that Seyfert galaxies have been spectroscopically monitored with sufficient frequency to make such detection possible. Whereas the continuum increased by a factor of ~ 1.7 , the He II line increased more than tenfold. The most likely explanation for this change was the addition of a new gaseous component. He II lines are formed by recombination in active nuclei⁶: the luminosity of a recombination line will remain constant if the gas is fully ionized and the mass and density of the emitting gas remain constant. In a radiation-bounded geometry, one in which the majority of the ionizing radiation striking a BLR cloud is absorbed by the emitting gas, the equivalent width of an emission line should remain constant, and the He II line should increase in phase with the continuum. Our observations rule out both these simple pictures.

It is possible to argue that the dramatic increase at 4,686 Å is caused by a flattening of the continuum, thereby increasing the number of helium-ionizing photons relative to the optical continuum. We reject such an explanation because our data show no evidence for continuum shape change at lower energies, based on ultraviolet and optical observations. Moreover, the

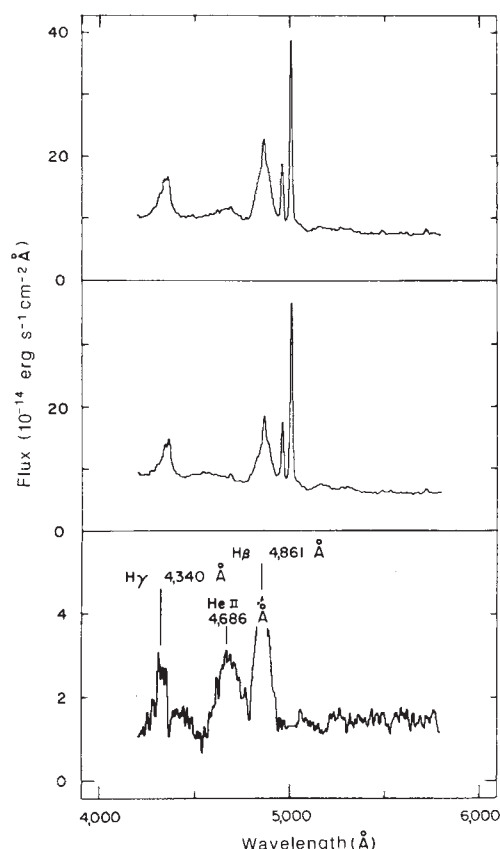


Fig. 2 The top two panels show the optical spectrum near the H β and He II 4,686 Å lines during the high and low states. The bottom panel is a difference spectrum, with the strongest emission lines marked. The high-state spectrum was obtained on 1–2 June 1984 UT (2,445,852–3 JD) and the low-state spectrum was obtained on 2–3 March 1984 UT (2,445,761–2 JD).

Balmer lines and the He I 5,876-Å line, which are sensitive to the radiation field in the range 1–4 Ryd, did not increase by as large a fraction as the optical continuum: if the increase at 4,686 Å is attributable to an increase in the density of ionizing photons, the increase in flux would have to be confined to energies in excess of 4 Ryd. We therefore believe that the increase in the $I(4,686 \text{ Å})/I(\text{H}\beta)$ ratio was caused by the appearance of a new component of gas.

The He II 4,686-Å line profile confirms that a new component of gas is present, that this gas is not associated with the classical BLR, and that the gas is probably closer to the central object than the BLR. Figure 2 shows a difference spectrum, obtained by subtracting a low-state spectrum (when He II emission is weak) from one obtained in the high state, when the anomalous He II emission was present. The increased H β line is apparent, along with H γ near the blueward limit of the spectrum. The H β profile is similar to that seen in low-state spectra, and is an indication of the width of broad emission lines in this object. Surprisingly, the new He II emission line has a width $\sim 75\%$ greater than the neighbouring H β . Although there is some tendency for He II lines to be slightly broader than hydrogen lines⁷, the difference seen here is much greater than has ever been seen before. The greatly enhanced He II emission can be attributed to gas moving with much greater velocity than BLR clouds. If the gas is confined within the potential well of the central object, we can infer that the He II-emitting material lies at roughly one-third of the BLR radius from the central object, or at roughly 3×10^{16} cm. We thus conclude that the increase in luminosity of the central object in NGC 5548 was accompanied by the appearance of a new component of gas much closer to the central object than previously detected. The coincidence further sug-

gests that the newly present very-broad-line-region (VBLR) gas may be associated with the increase in accretion rate that must occur in the standard picture.

At first sight, it seems surprising that the VBLR is detected only in He II. The lack of a correspondingly spectacular change in H β or He I 5,876 Å is convincing evidence that the gas is fully ionized, namely, He is doubly ionized throughout the VBLR. If so, then the $I(4,686 \text{ Å})/I(\text{H}\beta)$ intensity ratio would be near its case B value of ~ 1 for a solar composition. Strong VBLR H β emission is not expected. Although H β appears much sharper than He II 4,686 Å, tests show that a fairly substantial component of very broad H β could exist in the observed profile. Line profile synthesis shows that the $I(4,686 \text{ Å})/I(\text{H}\beta)$ ratio in the VBLR lies between 1 and 2. If both lines radiate with case B emissivities then this ratio sets the limits $0.1 \leq \text{He}/\text{H} \leq 0.2$, that is, the material does not seem to be highly processed.

A simple argument suggests that the new component of gas nearly covers the continuum source completely. For conditions in the BLR, where the gas is optically thick to ionizing radiation, the equivalent width of Ly- α is related to the mean optical-X-ray continuum slope by $W(\text{Ly-}\alpha) = (1,216/\alpha)(912/1,215)^\alpha (\Omega/4\pi)$, where α is the spectral index and $\Omega/4\pi$ is the covering factor of absorbing gas. This assumes that each ionizing photon eventually produces one Ly- α . The spectral index is $\alpha_{\text{ox}} \approx 1.1$, corresponding to a BLR covering factor of $\sim 15\%$. This value is fairly typical of AGN. The He II $I(4,686 \text{ Å})/I(\text{H}\beta)$ ratio before the increase in luminosity was also fairly typical of active nuclei (see Fig. 1). If we take the pre-increase 4,686-Å flux as typical of gas with a 15% covering factor, then the roughly sevenfold increase in equivalent width at 4,686 Å corresponds to an increase in the covering factor to 50–100%. The increased equivalent width further suggests that the VBLR gas absorbs the majority of the He⁺-ionizing radiation ($\lambda \leq 228 \text{ Å}$). This fact, together with the large covering factor, suggests that it is this ephemeral VBLR that may account for the variable low-energy X-ray absorption characteristic of low-luminosity Seyfert galaxies⁹. The fact that the VBLR does not contribute He I lines (He I lines are present in our spectra, but have widths similar to the Balmer lines) shows that the VBLR is not, however, optically thick to He⁺-ionizing radiation.

Simple calculations allow estimation of the column density and mass of the VBLR gas. For an assumed distance of 70 Mpc the 4,686-Å flux at maximum light is $\sim 1.0 \times 10^{41} \text{ erg s}^{-1}$. This corresponds to an emission measure of $N^2 V = 8 \times 10^{66} \text{ cm}^{-3}$, assuming a temperature of 10^5 K and $\text{He}/\text{H} = 0.1$, where N is the hydrogen density and V the emitting volume. There is no reliable density indicator in the VBLR spectrum that we observe, but if we assume a density of 10^{10} cm^{-3} (a typical BLR density) and full coverage at a distance of $3 \times 10^{16} \text{ cm}$ then the VBLR gas has a column density of $\sim 7 \times 10^{22} \text{ cm}^{-2}$, typical of X-ray absorbing gas. The geometry of this gas is unclear because unpublished photoionization calculations show that such clouds would be unstable to disruption by radiation pressure. The emission measure also corresponds to a mass of $0.8 M_\odot$, assuming a solar composition. An obvious possibility is that the rapid appearance of the VBLR was the result of the disruption of a star. Such conjectures are beyond the scope of this report, but a possibility is that a star was disrupted in the inner regions of the system when the optical continuum began to rise, but that the newly added gas only became spectroscopically active in the optical region when debris was ejected to nearly the BLR radius.

We have shown that a new emission-line region, not previously detected in the optical spectrum, appeared when the non-thermal continuum in NGC 5548 increased in luminosity. If current ideas about the energy source in AGN are correct, then such an increase in luminosity can only be caused by an increase in the accretion rate in the innermost regions of the nucleus. We believe that we have actually observed such an accretion event in progress.

Received 14 July; accepted 1 September 1986.

1. Lynden-Bell, D. *Nature* **223**, 690-694 (1969).
2. Rees, M. J. *Physica Scr.* **17**, 193-200 (1978).
3. Begelman, M. C. in *Astrophysics of Active Galaxies and Quasi-Stellar Objects* (ed. Miller, J.) 711-752 (University Science Books, Mill Valley, California, 1985).
4. Peterson, B. M. *Astrophys. J.* **312** (in the press).
5. Peterson, B. M. & Gaskell, C. M. *Astr. J.* **91**, 552-556 (1986).
6. MacAlpine, G. M., Davidson, K., Gull, T. R. & Wu, C. C. *Astrophys. J.* **294**, 147-152 (1985).
7. Osterbrock, D. E. & Shuder, J. *Astrophys. J. Suppl.* **49**, 149-174 (1982).
8. Mushotzky, R. F., Marshall, F., Boldt, E., Holt, S. & Serlemitsos, P. *Astrophys. J.* **235**, 377-385 (1980).
9. Reichert, G. A., Mushotzky, R. F., Petre, R. & Holt, S. S. *Astrophys. J.* **296**, 69-89 (1985).

Large-amplitude semidiurnal temperature variations in the polar mesopause: evidence of a pseudotide

R. L. Walterscheid*, G. G. Sivjee*,
G. Schubert* & R. M. Hamway†

* Space Sciences Laboratory, The Aerospace Corporation,
POB 92957, Los Angeles, California 90009, USA

† Department of Physics and Astronomy, Dartmouth College,
Hanover, New Hampshire 03755, USA

Current atmospheric tidal models^{1,2} predict an amplitude of ≤ 2 K for the solar semidiurnal temperature variations in the polar mesopause (~ 85 km altitude) during the winter solstice period. In contrast, here we report an amplitude of about 13 K for the 12-h variations in the mesopause temperature at Longyearbyen, Spitsbergen (78.2° N) on UT day 359 in 1984. The kinetic temperature of the mesopause was deduced from the thermalized rotational distribution of the mesopause airglow OH(8,3) vibrational-rotational band emission in the near-infrared spectral region. A semidiurnal 'pseudotide' is proposed as the source of the observed large-amplitude 12-h temperature variation. This pseudotide is generated by gravity-wave momentum fluxes modulated by the true semidiurnal tide³. The semidiurnal pseudotide is very efficiently excited at very high latitudes. Based on conservative values of gravity-wave amplitudes, we estimate a temperature amplitude of ~ 10 K.

Certain exothermic chemical reactions among mesopause constituents, particularly the interaction of O_3 with H, produce OH in vibrationally and rotationally excited states. The radiative lifetime of these radicals is ~ 1 ms, whereas the collision frequency in the mesopause is $\sim 10^4$ s⁻¹. Consequently the OH radicals undergo ~ 10 collisions, on average, before shedding their excess energy, and their distribution among the rotational states is thermalized to the ambient kinetic temperature. Measurements of the relative intensities of the airglow OH rotational-line emissions can therefore provide a measure of the mesopause kinetic temperature.

Spectroscopic observations of the OH rotational lines in the P_1 and P_2 branches of the (8,3) and the (6,2) vibrational-rotational bands were made in Longyearbyen, Spitsbergen, in December 1984. The high latitude of the observatory (78.2° N) permitted continuous 24-h observations of the wintertime polar airglow. On UT day 359 in 1984 the spectroscopic observations around 7,300 and 8,500 Å showed no auroral band emissions or O II 7,320-Å line emission, and only a weak O I 8,446-Å line emission. The absence of a significant level of any auroral optical emission indicates that there were no particle precipitation events during this period energetic enough to have had any effect on the thermodynamics of the mesopause. The local solar shadow height above the Longyearbyen observatory was > 150 km over the entire observing period.

We analysed the relative intensities of the spectrally resolved OH rotational lines, using Mies's⁴ transition probabilities and Kvifte's⁵ as well as Coxon and Foster's⁶ rotational-term values and emission wavelengths, to derive the mesopause temperature,

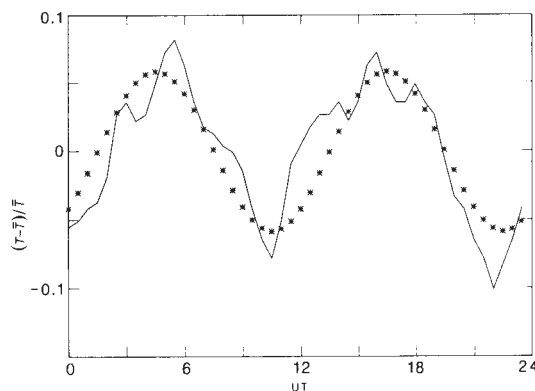


Fig. 1 Time history of normalized changes in mesopause temperature, that is, $(T - \bar{T})/\bar{T}$, for a 24-h period on UT day 359 in 1984 (continuous curve). The contribution attributable to the 12-h Fourier component (*) is also shown.

T . The average temperature, $\bar{T}$, of the mesopause over the 24-h period on UT day 359 in 1984 was 219 ± 2 K. For comparison with other reports of mesopause temperature variations (albeit of shorter period; see, for example, Krassovsky⁷), we have analysed the normalized changes in temperature, that is, $(T - \bar{T})/\bar{T}$. Figure 1 shows the time history of this quantity for the 24-h period on UT day 359 in 1984, and also the 12-h component of the temporal spectrum calculated by Fourier analysis of the temperature-data set. The major mode of the wave-like structure in the time history of $(T - \bar{T})/\bar{T}$ is clearly the semi-diurnal wave with an amplitude of 0.0574, which corresponds to $T - \bar{T} \approx 13$ K.

Forbes's² model calculations of the semi-diurnal components of the solar and lunar atmospheric tides predict amplitudes of 2.0 and 0.4 K at 87 km height in the polar region (78° N) during the December solstice period. Our calculations, using the model described in ref. 1, predict even smaller amplitudes (< 1 K). This model includes a large number (8) of the conventional zonal wavenumber 2 semidiurnal modes. Clearly, the normal tidal oscillations of the atmosphere do not account for the order-of-magnitude larger amplitude of the observed semidiurnal temperature oscillations. The observed temperature oscillations also cannot be due to very-high-order solar-driven modes, which potentially have significant amplitudes at high latitudes, because these modes are only weakly excited by global solar forcing, and local solar forcing is lacking in winter. The observed temperature oscillations might be due to the viscous distortion of the principal solar-driven modes. This distortion can broaden the latitude region having significant amplitude. However, such strong broadening has not been seen at emission altitudes in a model that includes this effect².

The occurrence of semidiurnal oscillations in polar-latitude mesopause temperatures is not an unusual event. Myrberg⁸ co-added 19 days of airglow data from Spitsbergen for the period 9-27 December 1982 and found a semidiurnal temperature amplitude of ~ 3 K, but no diurnal variation. Viereck *et al.* (R. A. Viereck, C. H. Deehr, G. G. Sivjee and G. J. Romick, manuscript to be submitted to *Planet Space Sci.*, hereafter referred to as VDSR) report that semidiurnal oscillations are seen in more than half the OH airglow data obtained from Spitsbergen during the winters of 1984-85 and 1985-86. In some of the data the semidiurnal oscillations are quite large; the co-addition of five consecutive days of data gave a predominantly semidiurnal variation with a peak-to-peak amplitude of ~ 16 K. Forbes², on the other hand, predicts a predominantly diurnal variation at the latitude of Spitsbergen. VDSR do report the presence of diurnal variations in their inferred mesopause temperatures.

The above considerations suggest the possibility of a high-latitude source for the oscillations such as Joule and particle heating in the lower thermosphere. However, conventional semidiurnal modes are not good candidates for exhibiting the effects

of high-latitude forcing because, even though extremely high-order modes can potentially have appreciable amplitude at high latitude, they are strongly absorbed by dissipative processes in the lower thermosphere and, moreover, are only weakly excited by high-latitude heating. Because these modes are not confined to high latitude, but are global in extent, high-latitude heating does not project well onto these modes. In fact, the totality of conventional semidiurnal modes (denoted gravitational modes by Flattery⁹) are not able to resolve very-high-latitude thermal forcing. Gravitational modes alone are not complete, and another set of modes called rotational modes⁹ are required when $\lambda = \omega/2\Omega < 1$, where ω is the wave angular frequency and Ω is the Earth's angular speed. For solar semidiurnal tides $\lambda = 0.9973$. Rotational modes with $\lambda \approx 1$ are trapped very close to the pole¹⁰. The main difficulty with invoking rotational modes as the source of the observed temperature variation is that these modes are evanescent and have very short vertical scales (≤ 300 m); they would therefore require very strong local forcing, which does not exist at mesopause altitudes during polar night. These modes might be viable if strong mean wind effects altered the effective rotation rate and if the OH airglow emissions were coming from altitudes much higher than now believed. Also significant in the data of Fig. 1 is the lack of any substantial power at the frequency of the diurnal tide. If the semidiurnal tide were responding to high-latitude forcing then the diurnal tide ought to be even more prominent because certain diurnal modes respond more efficiently to high-latitude sources than do the semidiurnal modes.

It is possible that the semidiurnal variation reflects near-inertial oscillations, because the inertial period at Spitsbergen is nearly semidiurnal (12.3 h). However, it seems unlikely that free inertial waves would show the phase coherence seen in Fig. 1 over two complete cycles. Moreover, inertial motions are mainly horizontal and are associated with small temperature fluctuations. Also, near-inertial waves have very short vertical wavelengths, but the time series for emission intensity for two different OH bands ((8,3) and (6,2)) show semidiurnal variations that are essentially in phase, although the emission heights for the two bands are separated by a few kilometres.

We propose that the observed semidiurnal temperature oscillation is a pseudotide driven by gravity-wave-mean-flow interactions modulated by the solar-driven tide. The pseudotide is generated as follows: the rate at which gravity waves force the mean flow increases as $|\bar{u} - c|$ decreases, where $\bar{u}$ is the component of the large-scale flow in the direction of wave propagation, and c is the phase velocity of the wave; both quantities are positive eastward. For convenience, consider zonally propagating two-dimensional waves with $\bar{u}$ the zonal component of the semidiurnal tide. Waves with $c > 0$ ($c < 0$) impart an eastward (westward) acceleration to $\bar{u}$. During the eastward (westward) phase of the semidiurnal oscillation eastward (westward) propagating waves are preferentially absorbed. The net effect is to induce a semidiurnal contribution to $\bar{u}$ which is non-tidal. This mechanism is discussed in more detail in ref. 3.

In the simple case of zonally propagating two-dimensional gravity waves the forcing of the mean flow can be expressed as $\partial^2 \bar{u} / \partial t^2 + f^2 \bar{u} = (\partial / \partial t) \bar{G}$, where $\bar{G} = -(1/\rho_0)(\partial / \partial z) \rho_0 \bar{u}' w'$, $f = 2\Omega \sin \phi$ is the Coriolis parameter (ϕ is latitude), ρ_0 is a reference density profile, z is height, and u' and w' are respectively the zonal and vertical gravity-wave velocities. The quantity $\bar{G}$ is the forcing of $\bar{u}$ due to the wave-momentum flux convergence. An overbar refers to a spatial average taken over a distance equal to the wavelength of the longest gravity waves making significant contributions to the momentum flux. When $\bar{G}$ is periodic with semidiurnal frequency $\omega \approx 2\Omega$, $\hat{u} = -[i\omega/(\omega^2 - f^2)] \hat{G} = -(i/\cos^2 \phi)(1/\omega) \hat{G}$, where $\hat{G}$, $\hat{u}$ are the Fourier transforms of $\bar{G}$, $\bar{u}$. At the latitude of Spitsbergen the amplification factor $\alpha = |\omega^2/(f^2 - \omega^2)|$ ($= 1/\cos^2 \phi$, for $\omega = 2\Omega$) is 23 for the semidiurnal frequency. By contrast, $\alpha = 0.35$ for the diurnal frequency. At 60° N, $\alpha = 4$ for the semidiurnal

frequency. Thus strong amplification in polar regions is limited to the semidiurnal frequency, and to very high latitude. A semidiurnal oscillation with an amplitude of $\sim 10 \text{ m s}^{-1}$ could be induced near mesopause heights for conservative gravity-wave amplitudes³ and $\alpha \sim 2$.

Thus it seems possible for a gravity-wave semidiurnal tide interaction at very high latitudes to force a large-amplitude pseudotide. The plausibility of the interaction is enhanced by the relatively large semidiurnal tidal-wind speeds. Semidiurnal tidal-wind velocities are not proportionately as small as temperature oscillations at high latitudes because the latitudinal structure functions for the wind components decay toward the poles much more slowly than do the temperature functions (that is, Hough functions). Our tidal model predicts semidiurnal tidal velocities of $\sim 20 \text{ m s}^{-1}$ near 90° km at 75° N during the winter. This estimate, together with the results presented in ref. 3 for a gravity-wave phase speed of $\sim 20 \text{ m s}^{-1}$, suggest pseudotide amplitudes of $\sim 50 \text{ m s}^{-1}$ at Spitsbergen, provided that momentum fluxes are comparable. In the calculations gravity-wave amplitudes were 10 m s^{-1} at 80 km , compared with observed r.m.s. amplitudes of $\sim 20\text{--}40 \text{ m s}^{-1}$ (refs 11, 12). Amplitudes at very high latitudes have not been measured, but the latitude dependence is apparently not strong¹¹⁻¹⁴. Wintertime observations at Heiss Island (80.4° N) show semidiurnal wind amplitudes of $\sim 10 \text{ m s}^{-1}$ (ref. 15). These values are averages over the meteor region ($\sim 80\text{--}100 \text{ km}$) and are monthly averages based on three winters. Because the dominant tidal modes at high latitude are apt to have short vertical wavelengths ($\sim 20 \text{ km}$) and because the induced tide is apt to be rather variable in time (reflecting variable gravity-wave fluxes), predictions of large semidiurnal velocities are consistent with the observations.

To estimate the induced oscillation of temperature we must take into account divergent motions that result when mean quantities vary in latitude. The temperature change, $\bar{T}$, resulting from adiabatic heating and cooling can be estimated as follows. From the continuity equation $\bar{w} \sim (M/L)\bar{u}$, where L is the meridional length scale and M is the vertical length scale. In obtaining this equation we have used $\bar{u} \sim \bar{v}$, where $\bar{v}$ is the meridional wind component. From the first law of thermodynamics $\bar{T} \sim (1/\omega)\Gamma_a[(M/L)\bar{u}]$, where the quantity in square brackets is $\bar{w}$ and where Γ_a is the adiabatic lapse rate ($= g/c_p$, where g is the acceleration of gravity and c_p is the specific heat of air at constant pressure). With $\Gamma_a = 10^{-2} \text{ K m}^{-1}$, $M \sim 20 \times 10^3 \text{ m}/2\pi \sim 3 \times 10^3 \text{ m}$, $L \sim 10^6 \text{ m}$ (L is the half width of the region over which α is large and the driving tide is sensible; the tidal amplitudes vanish at the pole) and $\bar{u} \sim 50 \text{ m s}^{-1}$, we obtain $\bar{T} \sim 10 \text{ K}$. This estimate agrees well with the observed amplitude of 13 K . It is also consistent with the results of Myrabø⁸ and VDSR mentioned above. These results were based on averages over a number of days because the day-to-day variability of the semidiurnal variation in temperature can be quite large. This large temperature variability perhaps reflects the unsteadiness of the gravity wave fluxes³.

We note finally that some authors have observed a connection between stratospheric warmings and changes in the amplitude of the semidiurnal tide (ref. 16 and VDSR). The observations presented here were made one day before the onset of a sudden warming was observed in stratospheric data¹⁷ and several days before the characteristic mesopause cooling was observed in airglow data (VDSR). Thus there is no obvious connection between the large-amplitude semidiurnal variation in the data we have presented and the sudden warming. However, VDSR observed an enhancement in semidiurnal amplitudes after the mesospheric cooling. Processes associated with sudden warmings can affect tides by redistributing ozone and modifying the thermal drive, or by changing the background wind system and modifying mode coupling^{18,19} or the transmission of vertically propagating gravity waves.

Neither thermal driving nor mode coupling seems viable because neither mechanism seems able to explain the

amplification of the very-high-order modes required to explain large-amplitude tides at very high latitudes. The required amplification of these modes cannot be due to ozone redistribution at polar latitudes because of the absence of sunlight, or to mode coupling because this process is ineffective in generating very-high-order modes from the lower-order energy-containing ones¹⁹. Moreover, there is no reason to believe that mode coupling is significantly stronger in the pre-warming atmosphere than in the post-warming one. In addition, it is very unlikely that changes in the thermal drive and mode coupling would act to amplify the semidiurnal tide and not the diurnal one. No such amplification of the diurnal oscillation is observed at Spitsbergen (VDSR).

The observed enhancement of the semidiurnal oscillation might be due to an enhanced flux of gravity waves associated with the breakdown of the polar vortex. An enhanced prevalence of large-amplitude waves has been observed (C. R. Philbrick, personal communication) and the semidiurnal pseudotide forced by this mechanism would be favoured over a one.

We thank the NSF for financial support under research grants ATM85-42125 and ATM84-12982. This work was also supported in part by the Aerospace Sponsored Research Program.

Received 23 June; accepted 1 October 1986.

- Walterscheid, R. L., DeVore, J. G. & Venkateswaran, S. V. *J. atmos. Sci.* **37**, 455-470 (1980).
- Forbes, J. M. *J. geophys. Res.* **87**, 5241-5252 (1982).
- Walterscheid, R. L. *J. geophys. Res.* **86**, 9698-9706 (1981).
- Mies, F. H. *J. molec. Spectrosc.* **53**, 150-188 (1974).
- Kvifte, G. *Geofis. Pubs* **20**, 1-17 (1959).
- Coxon, J. A. & Foster, S. C. *Can. J. Phys.* **60**, 41-48 (1982).
- Krassovsky, V. I. *Ann. Geophys.* **28**, 739-746 (1972).
- Myrback, H. K. *Planet. Space Sci.* **37**, 249-255 (1984).
- Flattery, T. W. thesis, Univ. Chicago (1967).
- Longuet-Higgins, M. S. *Phil. Trans. R. Soc. A262*, 511-607 (1968).
- Vincent, R. A. *J. atmos. terr. Phys.* **46**, 119-128 (1984).
- Mason, A. H., Gregory, J. B. & Stephenson, D. G. *J. atmos. Sci.* **32**, 1682-1688 (1975).
- Hirota, I. in *Dynamics of the Middle Atmosphere* (eds Holton, J. R. & Matsuno, T.) 65-75 (Terra Scientific, Tokyo, 1984).
- Vincent, R. A. *J. atmos. terr. Phys.* **46**, 961-974 (1984).
- Lysenko, I. A., Orlyanskii, A. D. & Portnyagin, Y. I. In *Coll. ext. Summ. Contrib. Joint Assembly Seattle, Washington, CMUA Sess. IAGA/IMAP, 22 August to 3 September 1977*, 31-1-31-7 (NCAR, Boulder, Colorado, 1977).
- Manson, A. H., Meek, C. E. & Labitzke, K. *Adv. Space Res.* **4** (4), 41-55 (1984).
- Quiroz, R. S. *J. geophys. Res.* **91** (in the press).
- Lindzen, R. S. & Hong, S.-S. *J. atmos. Sci.* **31**, 1421-1446 (1974).
- Walterscheid, R. L. & Venkateswaran, S. V. *J. atmos. Sci.* **36**, 1623-1635 (1979).

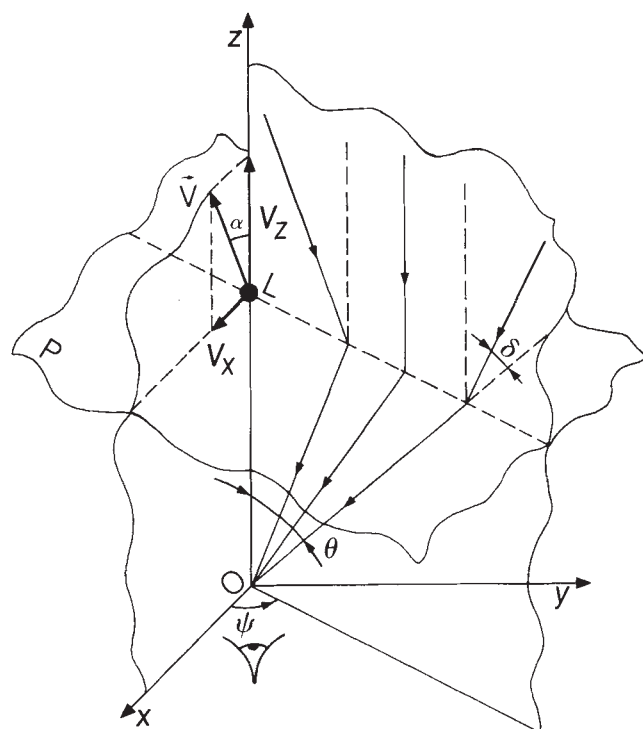


Fig. 1 The geometry of the problem: O is the observer, L is the position of the gravitational lens moving with velocity V in the X - Z plane, P is the picture plane, α is the angle between V and the Z axis, δ is the deflection angle, ψ is the position angle, and θ is the angle between the Z axis and the line of sight.

directions of propagation of the radiation at A and C, respectively. To find the background brightness after lensing by a moving mass one uses transformation (1) twice: the first time to transform from the background co-moving frame to the frame of a moving GL, and the second time to return from the moving GL frame to the background frame. The deflection of geodesic lines in the GL frame is a cause of the perturbation in background brightness.

Using this procedure one obtains for $\beta = V/c \ll 1$ the brightness perturbations

$$\frac{\Delta B(\nu, \mathbf{n})}{B_0(\nu)} = 2\beta [\cos \alpha \sin(\theta - \frac{1}{2}\delta) \sin \frac{1}{2}\delta - \sin \alpha \cos \psi \cos(\theta - \frac{1}{2}\delta) \sin \frac{1}{2}\delta] \times \frac{(h\nu/kT) \exp(h\nu/kT)}{\exp(h\nu/kT) - 1} \quad (2)$$

where B_0 and T are the initial background brightness; temperature and angles α , δ , ψ and θ are defined in Fig. 1.

In the Rayleigh-Jeans (RJ) part of a black-body spectrum the last factor of (2) is about 1. However, for the Wien part it is proportional to the frequency, but the brightness is exponentially small. In the square brackets the first item corresponds to the longitudinal motion of GL. It agrees with the previous result³ in the RJ spectral region. On the other hand, in ref. 3 the 'transverse' part of the effect, corresponding to the second item of (2), was ignored. For $\delta \ll \theta \ll 1$ the 'transverse' effect dominates the 'longitudinal' one. It should be noted that transverse and longitudinal effects form different patterns in the picture plane. The former does not depend on ψ , and thus the corresponding perturbation looks like a spot with circular symmetry; the latter includes the dipole term $\cos \psi$ and the corresponding perturbation has a central dipole asymmetry.

Perturbation of the background radiation by a moving gravitational lens

L. I. Gurvits & I. G. Mitrofanov

Space Research Institute, Academy of Sciences USSR, Profsojuznaja 84/32, 117810 Moscow, USSR

Birkinshaw and Gull¹ considered the perturbation of the cosmic background radiation by a moving gravitational lens (GL). An estimate of the maximum amplitude makes it seem possible that this effect may soon be observed, leading to the measurement of non-Hubble velocities of cosmological objects. Such measurements may become a significant test for cosmological models. We present here the general formula for this effect, which differs essentially from that in ref. 1.

Brightness $B(\nu, \mathbf{n})$ is transformed from a reference frame (A) to another (C), moving with velocity V_{AC} as follows:²

$$B_C(\nu, \mathbf{n}_C) = (1 - \Delta_{AC})^3 B_A[\nu/(1 - \Delta_{AC}), \mathbf{n}_A] \quad (1)$$

where

$$\Delta_{AC} = 1 - \frac{(1 - V_{AC}^2/c^2)^{1/2}}{1 - V_{AC} \cdot \mathbf{n}_C/c} \approx - \frac{V_{AC} \cdot \mathbf{n}_C}{c}$$

ν is the radiation frequency in frame C, and $\mathbf{n}_A$ and $\mathbf{n}_C$ are the

Comparing (2) with equation (12) in ref. 1 one may conclude that the latter takes into account only the transverse effect. What is more, there is a discrepancy by a factor of -2 between amplitudes of the effect for the RJ spectral region. It seems that equation (12) in ref. 1 was obtained by excluding the factor $(1 - \Delta_{AC})^3$ from the formula of brightness transformation such as (1).

An order of effect for the RJ spectrum may be estimated by a simple expression, which follows from (2),

$$\Delta B/B \sim \Delta T/T \sim \beta \delta \quad (3)$$

Recently Collins *et al.*⁴ found indications of non-Hubble velocities of galaxies of the order of $1,000 \text{ km s}^{-1}$, i.e. $\beta \approx 3 \times 10^{-3}$. One may use the angle between two lensed images to estimate the deflecting angle. There are some proposals (refs 5 and 6) that the two quasars 1146+111 B and C, separated by $157''$, are in fact two lensed images. This value corresponds to $\delta \approx 0.7 \times 10^{-3}$. Using these optimistic values we obtain $\Delta T/T \sim 2 \times 10^{-6}$. This is not far from today's sensitivity level, so the effect must be taken into account in future experiments.

We shall present elsewhere a general analysis of the complete effect based on this formula, both for $\delta < \theta$ and for $\delta \gg \theta$, and also for all β , not just for $\beta \ll 1$.

Received 13 May; accepted 22 August 1986.

1. Barkinshaw, M. & Gull, S. F. *Nature* **302**, 315-317 (1983).
2. Zel'dovich, Ya.B. & Novikov, I. D. *The Structure and Evolution of the Universe*. (University of Chicago Press, 1983).
3. Mitrofanov, I. G. *Pis'ma Astr. Zh. (Soviet Astr. Lett.)* **7**, 73-75 (1981).
4. Collins, C. A., Joseph, R. D. & Robertson, N. A. *Nature* **320**, 506-508 (1986).
5. Paczynski, B. *Nature* **319**, 567-568 (1986).
6. Turner, E. L. *et al. Nature* **321**, 142-144 (1986).

The near-parabolic flux and the origin of short-period comets

M. E. Bailey

Department of Astronomy, The University,
Manchester M13 9PL, UK

The origin of short-period comets, usually believed to arise by planetary perturbations acting on nearly parabolic orbits, has been unclear for many years¹. Given the observed near-parabolic flux, the predicted number of such comets is too small by a factor of ~ 100 (refs 2-15). It has been suggested, therefore, that the adopted near-parabolic flux is in error¹⁶, that a physical effect such as fragmentation^{3,17,18} could operate, or that short-period comets do not come from the outer Oort cloud at all^{9,10,13,19}. It has been proposed^{13-15,20,21} that the problem could be resolved by postulating a comet belt or massive inner core to the Oort Cloud. Here I recalculate the flux of nearly parabolic orbits from the Oort Cloud, including the recently discovered dominant injection of comets by the galactic tide²²⁻²⁵. The observed number of short-period comets is consistent with the Oort Cloud having a massive inner core²⁶.

Consider a spherically symmetrical comet cloud characterized by an energy distribution $N(E)$ defined²⁶ so that the number of comets with energy per unit mass E in the range $(E, E + dE)$ is

$$N(E) dE = K(E/\hat{E})^{-\gamma} dE \quad (1)$$

where γ is the index of the energy spectrum ($\gamma = 5/2$ for Oort's standard model²⁷), $E = -GM_{\odot}/2a$ is the specific energy of an orbit with semi-major axis a , $\hat{E} = -GM_{\odot}/2R_0$, $R_0 \approx 2 \times 10^5 \text{ AU}$ is the outer radius of the cloud and K is a constant defined²⁶ in terms of the observed near-parabolic flux and other parameters of the model. One can assume that comets with perihelion distances $q < q_{lc}$ are removed from the Oort Cloud by planetary

perturbations, and adopt²⁸ $q_{lc} \approx 15 \text{ AU}$. This defines a loss cylinder in velocity space, containing nearly parabolic comets with specific angular momenta $J < J_{lc} = (2GM_{\odot}q_{lc})^{1/2}$.

It is now possible to calculate the rate of injection of comets into the loss cylinder by stars and the galactic tide. The stellar injection rate per unit semi-major axis is^{24,26}

$$F_{lc,*}(a) = 4\pi^2 f(a) J_{lc}^2 \frac{GM_{\odot}}{2a^2} \begin{cases} \frac{\sigma_j^2(a)}{k^2(a) J_{lc}^2} & a < a_* \\ 1 & a > a_* \end{cases} \quad (2)$$

where a_* (see below) is the critical semi-major axis above which stellar perturbations alone are able to fill the loss cylinder, $f(a)$ is the phase-space distribution function defined in terms of equation (1) by²⁶

$$f(a) = \frac{K}{8\pi^3} R_0^{-\gamma} (GM_{\odot})^{-1/2} a^{\gamma-5/2} \quad (3)$$

and $\sigma_j(a)$ is the r.m.s. change in specific angular momentum per revolution due to stellar perturbations. $k(a)$ is a slowly varying logarithmic function on the order of 2, which can be taken²⁴ to be

$$k^2(a) = \frac{1}{2} [\ln(a/2q_{lc}) - 1] \quad (4)$$

and²⁸

$$\sigma_j^2(a) = \frac{4}{3} P(a) \langle r^2 \dot{\epsilon}_*(r) \rangle \quad (5)$$

where $P(a) = 2\pi(GM_{\odot})^{-1/2} a^{3/2}$ is the orbital period and $\dot{\epsilon}_*(r)$ is the mean stellar energy transfer rate to comets at r with semi-major axes a . (Heisler and Tremaine's expression²⁴ is not used because, as they remark, their result is invalid for $a \leq 4 \times 10^4 \text{ AU}$.) The mean stellar energy transfer rate is given approximately by²⁹

$$\dot{\epsilon}_*(a) \approx \frac{4\pi G^2 M_*^2 n_*}{V_*} \begin{cases} (a/a_c)^2 & a < a_c \\ 2 \ln(a/a_c) + 1 & a > a_c \end{cases} \quad (6)$$

where $a_c \equiv \sqrt{\frac{12}{7}} b_{\min}$ and $b_{\min} = (2\pi n_* V_* T)^{-1/2}$ is the smallest impact parameter expected during the time T over which $\dot{\epsilon}_*(a)$ is averaged. Thus,

$$\frac{\sigma_j^2(a)}{k^2(a) J_{lc}^2} = \frac{80\pi^2}{3} \frac{(GM_{\odot})^{1/2}}{q_{lc}} \frac{n_* M_*^2}{V_* M_{\odot}^2} a^{7/2} g(a) \quad (7)$$

where

$$g(a) = \frac{1}{[\ln(a/2q_{lc}) - 1]} \begin{cases} (a/a_c)^2 & a < a_c \\ 2 \ln(a/a_c) + 1 & a > a_c \end{cases} \quad (8)$$

This allows the stellar-induced flux into the loss cylinder and a_* to be evaluated. Being concerned with filling the loss cylinder during one revolution, it is necessary to set $T = P(a)$, and adopt $n_* = 0.1 \text{ pc}^{-3}$, $V_* = 30 \text{ km s}^{-1}$ and $\langle n_* M_*^2 \rangle = \frac{4}{3} \times 0.03 M_{\odot}^2 \text{ pc}^{-3}$, where the factor $\frac{4}{3}$ is to include the effects of close binaries³⁰. This leads to $a_* \approx 5.0 \times 10^4 (q_{lc}/15 \text{ AU})^{2/7} \text{ AU}$, significantly larger than the value previously obtained²⁴. This strengthens the conclusion^{23,24} that stellar perturbations are relatively ineffective at filling the loss cylinder. (The results are different because with $T = P(a)$, $b_{\min}$ is typically in the range $1-2 \times 10^4 \text{ AU}$ and $g(a) \leq 1/2$. For comparison, if one sets $T = T_{ss} = 4.5 \times 10^9 \text{ yr}$, then $b_{\min} \approx 700 \text{ pc} \ll a$ and $g(a) \geq 1$. a_* is then $3.6 \times 10^4 (q_{lc}/15 \text{ AU})^{2/7} \text{ AU}$, in agreement with the earlier value²⁴.) A similar result for a_* has been independently obtained by Fernández and Ip³¹.

Next, consider the injection rate of comets by the galactic tide. This reduces²⁴ to

$$F_{lc,t}(a) = 4\pi^2 f(a) J_{lc}^2 \frac{GM_{\odot}}{2a^2} \begin{cases} \frac{10}{3\sqrt{2}} \frac{C_{gal}}{GM_{\odot} q_{lc}^{1/2}} a^{7/2} & a < a_t \\ 1 & a > a_t \end{cases} \quad (9)$$

where a_t , the critical semi-major axis above which the tide just fills the loss cylinder, is given by

$$a_t = \left[\frac{3\sqrt{2}}{10} GM_\odot q_{lc}^{1/2} C_{gal}^{-1} \right]^{2/7} \approx 3.3 \times 10^4 \left(\frac{q_{lc}}{15 \text{ AU}} \right)^{1/7} \text{ AU} \quad (10)$$

Here a galactic force law perpendicular to the galactic plane of the form $g_z = -C_{gal}z$, with $C_{gal} \approx 10^{-29} \text{ s}^{-2}$ has been assumed³². This corresponds to a mid-plane density (see refs 24, 33) $\rho_0 = C_{gal}/4\pi G = 0.176 M_\odot \text{ pc}^{-3}$. For $a < a_t$ the ratio of the tidal to stellar injection rates is therefore

$$R_{t/*} = \frac{\sqrt{2}}{16\pi^2} \frac{q_{lc}^{1/2} V_*}{(GM_\odot)^{3/2}} \frac{C_{gal}}{1} \frac{1}{M_\odot^2 \langle n_* M_*^2 \rangle g(a)} \quad (11)$$

which, with the adopted values of the parameters, reduces to $R_{t/*} = 1.933/g(a)$. Because $g(a) \approx 1/5$ for $a = a_c \approx 2 \times 10^4 \text{ AU}$ one can conclude that the galactic tide dominates the stellar injection rate by about an order of magnitude at this value of a . For smaller a values, the tidal dominance of the injection rate increases.

This result (and that for a_* above) relies on equation (2), which in turn depends on a diffusion approximation to describe the quasi-steady distribution of comets near the loss cylinder. For $a \leq 2 \times 10^4 \text{ AU}$ this assumption begins to break down because $b_{\min}$ becomes greater than a . Thus, although this approach should correctly give the general trend, the results are not expected to be precise. A recent Monte Carlo simulation by Heisler *et al.*³⁴ indicates that the galactic tide dominates the stellar injection rate by a factor of ~ 4 at $a = 2 \times 10^4 \text{ AU}$.

An approximate total injection rate may thus be obtained by considering only the galactic tide. If a_0 is the semi-major axis corresponding to the inner edge of the Oort cloud, then

$$\dot{N}_{\text{tot}} \approx \dot{N}_{\text{tot},t} = \int_{a_0}^{R_0/2} F_{lc,t}(a) da \quad (12)$$

Substituting for f from (3) and introducing $x = a/a_t$, $\alpha_R = R_0/2a_t \gg 1$ and $\alpha_0 = a_0/a_t \ll 1$, this reduces to

$$\dot{N}_{\text{tot}} = \frac{KR_0^{-\gamma}}{2\pi} (GM_\odot)^{3/2} q_{lc} a_t^{\gamma-7/2} \left\{ \int_1^{\alpha_R} x^{\gamma-9/2} dx + \int_{\alpha_0}^1 x^{\gamma-1} dx \right\} \quad (13)$$

In this expression, the first term corresponds to comets with almost parabolic orbits and a nearly uniform distribution of perihelion distances, whereas the second corresponds to comets with rather shorter periods and perihelia close to q_{lc} (for example ref. 28). The comets in this second group cannot be directly observed (as $q_{lc} \approx 15 \text{ AU}$), but those in the first group can be detected up to distances of the order of 5 AU. This allows the leading term of (13) to be re-written in terms of the observed nearly parabolic flux F_{obs} , defined here as the number of comets passing perihelion per unit time with perihelia less than some value D_{obs} . The ratio $F_{\text{obs}}/D_{\text{obs}}$ is then nearly constant. Thus, from (13),

$$F_{\text{obs}} = \frac{KR_0^{-\gamma}}{2\pi} (GM_\odot)^{3/2} q_{lc} a_t^{\gamma-7/2} \int_1^{\alpha_R} x^{\gamma-9/2} dx \frac{D_{\text{obs}}}{q_{lc}} \quad (14)$$

so

$$\dot{N}_{\text{tot}} = \frac{F_{\text{obs}}}{D_{\text{obs}}} q_{lc} \left\{ 1 + \frac{\int_{\alpha_0}^1 x^{\gamma-1} dx}{\int_1^{\alpha_R} x^{\gamma-9/2} dx} \right\} \quad (15)$$

The total number of short-period comets may now be obtained¹³⁻¹⁵ by multiplying $\dot{N}_{\text{tot}}$ by three factors: the fraction f_0 of comets in the loss cylinder with small enough inclinations ($0 \leq i \leq i_0$) to be significantly affected by planetary perturbations; the mean planetary capture probability for each group of incoming comets including the effects of all the planets; and the average lifetime T_{sp} of observable short-period comets. The

first factor may be approximately estimated by assuming that the incoming comets have random inclinations. In this case, although the assumption is not strictly accurate when the galactic tide drives the influx, $f_0 = \frac{1}{2}(1 - \cos i_0)$. The second factor may be written as an integral over perihelion distance of the mean planetary capture probability $p(q)$, weighted according to the perihelion distribution of each group of comets. For the first group (distributed uniformly with q) this becomes

$$\bar{p} = \frac{1}{q_{lc}} \int_0^{q_{lc}} p(q) dq \quad (16)$$

whereas for the second group (concentrated close to $q = q_{lc}$) it is simply $\bar{p} = p(q_{lc})$. Following Fernández¹³ in assuming that the mean planetary capture probability $p(q)$ has the step-wise form

$$p(q) = \begin{cases} 1/128 & 0.0 < q \leq 5.8 \text{ AU} \\ 1/700 & 5.8 < q \leq 10.7 \text{ AU} \\ 1/3,200 & 10.7 < q \leq 22 \text{ AU} \\ 1/6,000 & 22 < q \leq 34 \text{ AU} \end{cases} \quad (17)$$

gives $\bar{p} \approx 0.054 \text{ AU}/q_{lc}$ for the first group, and $\bar{p} = 1/3,200$ or $1/700$ for the second group, depending respectively on whether the planetary captures are dominated by Uranus ($q_{lc} = 15 \text{ AU}$) or Saturn ($q_{lc} = 10 \text{ AU}$). Adopting the values for Uranus, and setting^{14,15} $i_0 = 30^\circ$, $T_{\text{sp}} = 3,000 \text{ yr}$, and $F_{\text{obs}} = 0.7 \text{ yr}^{-1}$, corresponding to $D_{\text{obs}} = 2.5 \text{ AU}$, then

$$N_{\text{sp}} = \frac{F_{\text{obs}}}{D_{\text{obs}}} T_{\text{sp}} \left\{ \int_0^{q_{lc}} p(q) dq + p(q_{lc}) q_{lc} \frac{\int_{\alpha_0}^1 x^{\gamma-1} dx}{\int_1^{\alpha_R} x^{\gamma-9/2} dx} \right\} \quad (18)$$

or

$$N_{\text{sp}} \approx 56.3 \left\{ 0.054 + \frac{15}{3200} F(\gamma) \right\} \quad (19)$$

where $F(\gamma)$ is the ratio of the integrals appearing in the second term of (18). The observed number of short-period comets ($\sim 10^2$) can thus be understood if $F(\gamma) \approx 200$. This is clearly not possible in Oort's standard model with $\gamma = 5/2$; but for arguably more plausible models^{26,30,35} with $\gamma < 0$ it is certainly possible. For example, adopting $R_0 = 2 \times 10^5 \text{ AU}$, $a_t = 3 \times 10^4 \text{ AU}$, $a_0 = 4 \times 10^3 \text{ AU}$ and $\gamma = -2$, then we find $F(\gamma) \approx 152$ and $N_{\text{sp}} \approx 43$. On this model the flux of long-period comets through the Uranus capture zone is more than two orders of magnitude greater than expected from a simple linear extrapolation of the observed local flux.

A check on the model is provided by the total number of comets. This is readily calculated²⁶ as

$$N_0 = KR_0^{-\gamma} \frac{GM_\odot}{2} \int_{a_0}^{R_0/2} a^{\gamma-2} da$$

that is,

$$N_0 = \frac{F_{\text{obs}}}{D_{\text{obs}}} \pi (GM_\odot)^{-1/2} a_t^{5/2} \frac{\int_{\alpha_0}^{\alpha_R} x^{\gamma-2} dx}{\int_1^{\alpha_R} x^{\gamma-9/2} dx} \quad (20)$$

For the above parameters this gives $N_0 = 1.7 \times 10^{13}$. With a mean mass per comet³⁶ $\sim 2.6 \times 10^{14} \text{ kg}$, the total mass of the cloud becomes $\approx 700 M_\oplus$, a large but not unreasonable amount³⁷. It is possible that such a mass of comets might already have been detected by the Infra-Red Astronomy Satellite³⁸⁻⁴⁰.

Thus I have shown that the observed number of short-period comets may be understood in terms of the capture of Oort Cloud comets by planetary perturbations, provided that the cloud has a massive inner core and that the injection of new comets is dominated by the galactic tide. This is because the ratio of the tidal injection rate to that of the stars increases sharply ($\propto a^2$) for semimajor axes $\leq 3 \times 10^4 \text{ AU}$, and although injection by the galactic tide is less effective at smaller semi-major axes (equation (9)), the relative decline may be more than off-set if there is a sufficiently large number of comets in the inner core. As with

some previous suggestions for overcoming the short-period comet problem^{9,10,14,19}, this analysis predicts that most short-period comets are initially captured by Uranus (or possibly Saturn), rather than Jupiter. In particular, the present theory predicts that most short-period comet progenitors are injected by the galactic tide with relatively small semi-major axes close to a_0 ($\approx 5,000$ AU). In this case the short-period comet population would bear little or no relation to the observed near-parabolic flux, and the usual links between these comets and the more numerous observed long-period group are brought into question. These results indicate the importance of detailed Monte Carlo simulations⁴¹ aimed at discovering whether such original orbits evolve into observable short-period comets.

I thank S. V. M. Clube and W. M. Napier for comments on the manuscript and S. D. Tremaine for helpful suggestions and a preprint³⁴. This work was supported by the SERC.

Received 31 July; accepted 1 October 1986.

1. Bailey, M. E., Clube, S. V. M. & Napier, W. M. *Vistas in Astr.* **29**, 53–112 (1986).
2. Proctor, R. A. *Knowledge* **6**, 126–128 (1884).
3. Bredichin, T. *Ann. de l'Obs. de Moscou* (Ser. 2) **2**, 1–17 (1889).
4. Russell, H. N. *Astr. J.* **33**, 49–61 (1920).
5. Proctor, M. & Crommelin, A. C. D. *Comets* (Technical Press, London, 1937).
6. Everhart, E. *Astr. J.* **74**, 735–750 (1969).
7. Havnes, O. *Icarus* **12**, 331–337 (1970).
8. Everhart, E. *Astrophys. Lett.* **10**, 131–135 (1972).
9. Vaghi, S. *Astr. Astrophys.* **24**, 107–110 (1973).
10. Joss, P. C. *Astr. Astrophys.* **25**, 271–273 (1973).

11. Kazimirschak-Polonskaya, E. I. *IAU Symp.* **45**, 373–397 (1972).
12. Everhart, E. *IAU Colloq.* **39**, 99–104 (1977).
13. Fernández, J. A. *Mon. Not. R. astr. Soc.* **192**, 481–491 (1980).
14. Fernández, J. A. & Ip, W.-H. *Asteroids, Comets, Meteors* (eds Lagerkvist, C.-I. & Rickman, H.) 387–390 (Uppsala Observatory, Uppsala, 1983).
15. Fernández, J. A. & Ip, W.-H. *Icarus* **54**, 377–387 (1983).
16. Hughes, D. W. *IAU Colloq.* **83**, 129–142 (1985).
17. Clube, S. V. M. & Napier, W. M. *Quart. Jl R. astr. Soc.* **23**, 45–66 (1982).
18. Napier, W. M. *Asteroids, Comets, Meteors* (eds Lagerkvist, C.-I. & Rickman, H.) 391–395 (Uppsala Observatory, Uppsala, 1983).
19. Delsemme, A. H. *Astr. Astrophys.* **29**, 377–381 (1973).
20. Mendis, D. A. *Astrophys. Sp. Sci.* **20**, 165–176 (1973).
21. Bailey, M. E. *Asteroids, Comets, Meteors* (eds Lagerkvist, C.-I. & Rickman, H.) 383–386 (Uppsala Observatory, Uppsala, 1983).
22. Torbett, M. V. *The Galaxy and the Solar System* (eds Matthews, M., Smoluchowski, R. & Bahcall, J. N.) (University of Arizona Press, Tucson, in the press).
23. Torbett, M. V. *Mon. Not. R. astr. Soc.* (in the press).
24. Heisler, J. & Tremaine, S. D. *Icarus* **65**, 13–26 (1986).
25. Morris, D. E. & Muller, R. A. *Icarus* **65**, 1–12 (1986).
26. Bailey, M. E. *Mon. Not. R. astr. Soc.* **204**, 603–633 (1983).
27. Oort, J. H. *Bull. Astr. Inst. Neth.* **11**, 91–110 (1950).
28. Bailey, M. E. *Mon. Not. R. astr. Soc.* **211**, 347–368 (1984).
29. Bailey, M. E. *Asteroids, Comets, Meteors II* (eds Lagerkvist, C.-I., Lindblad, B. A., Lundstedt, H. & Rickman, H.) 207–210 (Uppsala University, Uppsala, 1986).
30. Bailey, M. E. *Mon. Not. R. astr. Soc.* **218**, 1–30 (1986).
31. Fernández, J. A. & Ip, W.-H. Preprint.
32. Wielen, R. *Astr. Astrophys.* **25**, 285–297 (1973).
33. Bahcall, J. N. & Bahcall, S. *Nature* **316**, 706–709 (1985).
34. Heisler, J., Tremaine, S. D. & Alcock, C. Preprint.
35. Hills, J. G. *Astr. J.* **86**, 1730–1740 (1981).
36. Fernández, J. A. *Astr. J.* **87**, 1318–1332 (1982).
37. Bailey, M. E. *Nature* **302**, 399–400 (1983).
38. Rowan-Robinson, M. *Mon. Not. R. astr. Soc.* **219**, 737–749 (1986).
39. Bailey, M. E. *Mon. Not. R. astr. Soc.* **205**, 42P–25P (1983).
40. Bailey, M. E., McBreen, B. & Ray, T. P. *Mon. Not. R. astr. Soc.* **209**, 881–890 (1984); *Corrigendum*: **211**, 255.
41. Rickman, H. & Vaghi, S. *Astr. Astrophys.* **51**, 327–342 (1976).

Inferences on the evidence for radioactive ⁵³Mn in the early Solar System

Typhoon Lee

Institute of Earth Sciences, Academia Sinica, PO Box 23–59, Taipei, Taiwan (R.O.C.) 10769

Several now-extinct radionuclides are believed to have been present in the early Solar System¹ and have been used to infer the time-scales for various processes during its formation. However, the precise interpretation of the chronometric significance of an extinct nuclide depends on whether isotopic heterogeneities exist in that nuclide and/or in its daughter product. This is probable because such heterogeneities have already been found in many elements. Birck and Allègre² have recently reported data suggesting the presence of a new extinct nuclide, ⁵³Mn (mean life $\tau = 5$ Myr). In this letter I show that their data are inconsistent with the prediction of a single-stage model in which both Mn and its daughter element Cr were isotopically homogeneous. I also explore the virtues and consequences of the three alternative interpretations of the inconsistency: (1) the host material has experienced multiple-stage evolution, (2) ⁵³Mn was heterogeneously distributed, or (3) ⁵³Cr has intrinsic anomalies.

The discovery of ²⁶Al, whose mean life is shorter than that of ⁵³Mn, in the early Solar System suggests the possibility that ⁵³Mn may have also existed³. Recent searches for ⁵³Mn began with that of Nitoh *et al.*⁴. Lee and Tera^{5,6} placed more stringent upper limits on its abundance. However, positive results were not obtained until Birck and Allègre² developed superior analytical techniques enabling high-precision isotopic measurement of Cr (ref. 7). They also discovered isotopic anomalies in ⁵⁴Cr in ordinary Allende inclusions, which were confirmed by Papanastassiou *et al.*⁸. Papanastassiou⁹ further discovered much larger ⁵⁴Cr and ⁵³Cr anomalies in FUN (fractionation and unknown nuclear) inclusions. The discovery of anomalies intrinsic to the Cr isotopes accentuates the importance of considering the effects of Mn and Cr isotopic heterogeneities in the discussion of the reported evidence for the presence of ⁵³Mn.

Figure 1a shows the systematics of ⁵³Mn–⁵³Cr evolution in a diagram of ⁵³Cr/⁵²Cr against ⁵⁵Mn/⁵²Cr. The time (before present) evolution of ⁵³Cr/⁵²Cr due to ⁵³Mn decay for object P formed at t_1 from source S can be described by

$$(^{53}\text{Cr}/^{52}\text{Cr})_P(t) = I_1 + (^{53}\text{Mn}/^{55}\text{Mn})(t_1)(1 - \exp(-(t - t_1)/\tau)) \times (^{55}\text{Mn}/^{52}\text{Cr})_P \quad (1)$$

provided that this object and its source share the same initial ⁵³Cr/⁵²Cr(t_1) ($\equiv I_1$) and ⁵³Mn/⁵⁵Mn(t_1) and that the object is a closed system against Mn and Cr transport. Notice that equation (1) represents a straight line in Fig. 1a connecting the object, the source and the initial, whose slope increases with time up to the initial Mn ratio, whereas its intercept remains fixed at the initial Cr ratio. However, at any time, the line always passes through the source composition for that time (heavy lines in Fig. 1a). A group of such objects formed simultaneously at t_1 with initially homogeneous Mn and Cr ratios will thus form a linear array. This is the well-known isochron for an extinct nuclide. If such a group, after forming from the source, has remained closed until now (that is, experienced only single-stage evolution) then an isochron would be observed today and taken as evidence for the presence of ⁵³Mn at the formation time of the group. I have thus shown that, in a single stage evolution model with initial isotopic homogeneities, the isochron always passes through the source point. Applying this to the Allende meteorite and considering either different inclusions or different minerals in a single inclusion as a group formed directly from the solar nebula, we would thus expect that all such isochrons should pass through the normal Solar System point.

Birck and Allègre² reported two 'isochrons', one of which consists of whole-rock data for the Allende inclusion with a slope corresponding to an initial ⁵³Mn/⁵⁵Mn ratio of 6.7×10^{-5} . The other consists of whole rock data for Allende inclusions with a slope corresponding to an initial ⁵³Mn/⁵⁵Mn ratio of 3.7×10^{-5} . Figure 1b shows these isochrons and also some key data points. The whole inclusion isochron hinges on the data for G6 as noted in ref. 2; also, it does not pass through SB. The mineral isochron data span a relatively narrow range in Mn/Cr, giving a difference in ⁵³Cr/⁵²Cr only six times the standard error. Further Cr isotopic studies, especially on primitive high-Mn/Cr samples, appear necessary to substantiate these ⁵³Mn isochrons.

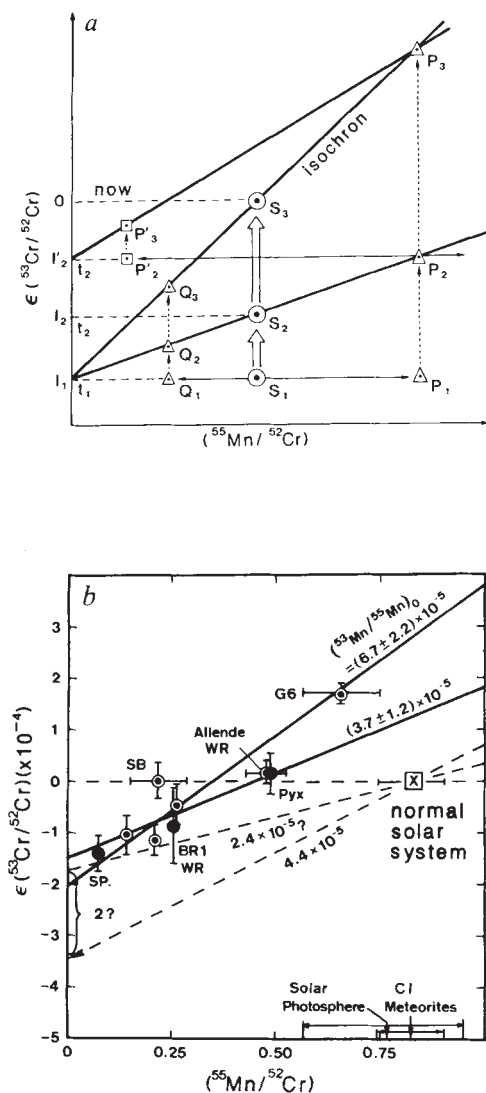


Fig. 1 Mn-Cr isochron diagram. $\epsilon(^{53}\text{Cr}/^{52}\text{Cr})$ is the fractional deviation of $^{53}\text{Cr}/^{52}\text{Cr}$ from the normal in unit of 0.01%. *a*, Schematics. The evolution of the Cr ratios by ^{55}Mn decay for a group of objects (P, Q , etc) formed from source S at time t_1 is described by equation (1) if they share the same initial Cr and Mn isotopic ratios as the source at t_1 and if they remain a close system. They plot as a linear array ('isochron') whose intercept is fixed at the initial Cr ratios I_1 and whose slope gradually steepens, eventually reaching the initial $^{55}\text{Mn}/^{52}\text{Mn}$ ratio. At any time the isochron for a group directly formed from the source (single-stage evolution) always passes through the source composition at that time, so the tie lines I_1S_i always coincide with isochrons P_iQ_i for $i = 1, 2, 3$ at time t_1, t_2 and 'now'. Therefore, in this single-stage homogeneous model, the isochrons for Allende inclusions are expected to pass through the putative source composition, that of the Solar System ($\odot$). In contrast, the Solar-System point may show an offset from the isochrons in a multiple-stage model. For example, the isochron for a group of objects like P' , which formed from an intermediate source reservoir P_2 at time t_2 that ultimately formed from S at an earlier time, does not pass through S . The direction of the offset depends on whether the Mn/Cr for P is higher or lower than that of S . *b*, Observed isochrons² and interpretations. The observed isochrons do not pass through the Solar-System point and the uncertainty in the Solar-System Mn/Cr estimates based on either meteoritic or photospheric observations does not seem to be large enough to eliminate the offset. If a two-stage evolution model for the Allende samples is invoked as the explanation, then the intermediate source reservoir must have a Mn/Cr higher than the solar value because of the direction of the offset. Alternatively, Mn isotopic heterogeneity can explain the offset if the solar reservoir has a $^{55}\text{Mn}/^{55}\text{Mn} = 2.4 \times 10^{-5}$ assuming that it and the samples started with the same initial $^{53}\text{Cr}/^{52}\text{Cr}$. The offset can also be explained if the initial $^{53}\text{Cr}/^{52}\text{Cr}$ for the Allende inclusions is higher than that of the normal solar reservoir by about 0.02%. The ratio of the inferred intrinsic ^{53}Cr excess to the observed ^{54}Cr excess (~ 0.3) for these ordinary inclusions is identical to that measured in the unusual FUN inclusion EK1-4-1 (ref. 9) and thus may have come from a single anomalous component of Cr.

In Fig. 1*b* I have also plotted the normal Solar System point at $\epsilon(^{53}\text{Cr}/^{52}\text{Cr}) = 0$ and $(^{55}\text{Mn}/^{52}\text{Cr}) = 0.85$ (ref. 10). The 'isochrons' do not pass through the Solar System points, which contrasts with the prediction of the above model.

We must first consider the possibility that the currently accepted Mn/Cr ratio for the Solar System is in error. The spectroscopically determined solar photospheric value is 0.60 (corresponding to a $^{55}\text{Mn}/^{52}\text{Cr}$ ratio of 0.76) with an uncertainty of $\sim 25\%$ (ref. 11). This estimate is consistent with the 'cosmic' value of 0.71, based on type CI meteorites, which has a nominal uncertainty of $\sim 13\%$ (see ref. 10 and references therein). However, this is based almost exclusively on the data of Orgueil, whereas some of the limited data on Alais are discrepant. An Mn/Cr of 0.3 is required to shift the Solar System point to the isochron. If this is the correct interpretation, our present estimate for the Solar System Mn/Cr has to be too high by about a factor of two. This seems unlikely in view of the consistency between the two estimates using different methods and their associated errors.

If the isotopic measurements and the estimated solar Mn/Cr value are correct then one or more of the three alternatives mentioned earlier must explain the offset of the Solar System point. Although the combination of more than one cause is possible it would be too intractable to allow a useful discussion, so I shall exclude the simple model's assumptions only one at a time. The first alternative, multiple-stage evolution of the host material, is difficult to quantify. This is because of the lack of constraints on the evolution during intermediate stages. There are many combinations of the duration and the Mn/Cr ratio for the intermediate stage that can explain the reported offset. Figure 1*a* shows a relatively simple two-stage evolution model that can be used to explain the offset and its direction. Object P_1 formed at t_1 from S_1 would reach point P_2 at t_2 ; hence a sample formed from P_2 , acting now as the intermediate source reservoir, would have the initial ratio I'_2 , much higher than I_2 , which is the initial for an object formed directly from S_2 . Connecting the initial with the intermediate source, line I'_2P_3 is the isochron observed today for the samples that experienced the two-stage evolution. It passes through P'_3 but not through S_3 , thus producing an offset from the Solar System point. We note that the direction of the offset depends on whether the intermediate-stage reservoir has a Mn/Cr ratio higher or lower than the solar value. Since the reported isochrons are to the left of the Solar System point, the intermediate-stage reservoir must have a Mn/Cr ratio higher than the solar value. This is somewhat surprising because in refractory samples such as Allende inclusions this ratio is lower than the solar value. The required high-Mn/Cr intermediate source reservoir from which the inclusions formed must be less refractory than the inclusions.

The remaining two alternatives have to do with isotopic heterogeneities rather than chemical fractionation. The offset of the normal Solar System point from the isochrons can be understood if Allende inclusions have a higher $^{53}\text{Mn}/^{55}\text{Mn}$ than the Solar System. However, because the initial $^{53}\text{Cr}/^{52}\text{Cr}$ of the inclusions is lower than normal and we assume that the initial Cr ratios of the inclusions were the same as the source (that is the solar) reservoir, the ^{53}Mn abundance in the normal Solar System material must be non-zero so that its decay can eventually increase the $^{53}\text{Cr}/^{52}\text{Cr}$ to the present value. The required $^{53}\text{Mn}/^{55}\text{Mn}$ ratio can be obtained by connecting the Solar System point to the initial Cr ratio of the observed isochrons. The slope of this line corresponds to a $^{53}\text{Mn}/^{55}\text{Mn}$ of about 2.4×10^{-5} (see Fig. 1*b*). We note that inclusion SB has a normal $^{53}\text{Cr}/^{52}\text{Cr}$ today. If it also started from the same initial Cr ratio as other inclusions then it coincidentally started from the right amount of ^{53}Mn , higher than that for other inclusions, to give a normal $^{53}\text{Cr}/^{52}\text{Cr}$ today. Or, more likely, it equilibrated later with the Solar System reservoir after all the ^{53}Mn had decayed away. This situation is reminiscent of the case of ^{26}Al where heterogeneities were also suspected¹.

Table 1 Estimated composition of the anomalous component

Nuclide	Observed abundance*	e-process abundance†		Normal
		$\eta = 0.0710$	$\eta = 0.0837$	
^{50}Cr	0	0.5	0.00001	1.8
^{52}Cr	0	419	2.7	36
^{53}Cr	1.3	1.2	0.18	4
^{54}Cr	1	1	1	1
^{53}Mn	0.7	16	0.003	—

* For assumptions see text.

† Ref. 12; all ^{54}Mn are assumed to have decayed into ^{54}Cr .

The last alternative is that the Allende samples began with a $^{53}\text{Cr}/^{52}\text{Cr}$ higher than that for the normal solar system material. The required amount can be found from Fig. 1b. If we draw an isochron through the Solar System point with a slope equal to the average of the published isochrons, this line would intercept the vertical axis at $\sim 0.02\%$ below the initial for the Allende samples. Thus the offset can be explained if the inclusions started with an initial intrinsic ^{53}Cr excess of 0.02% relative to the Solar System ratio at the time of formation of the inclusions. Because these samples already have excesses in ^{54}Cr of about 0.06% ^{7,8}, it seems plausible that the neighbouring neutron-rich Cr isotope also has some accompanying excesses. The possibility that part of ^{53}Cr effects in ordinary inclusions is intrinsic to this nuclide is enhanced by the recent discovery of a large ^{53}Cr excess in FUN inclusion EK1-4-1 by Papanastassiou⁹, who concluded that it is more reasonable to interpret the large effect as the result of intrinsic variations in ^{53}Cr than by invoking the *in situ* decay of ^{53}Mn . If so, and if both of the observed ^{53}Cr and ^{54}Cr effects in EK1-4-1 are caused by mixing an anomalous component with the normal, then the ratio of ^{53}Cr to ^{54}Cr in that component would be $0.16\%/0.48\% = 0.3$ times the normal ratio. After allowance is made for the contribution of ^{53}Mn effects, the ratio of the inferred intrinsic ^{53}Cr excess to the observed ^{54}Cr excess in ordinary inclusions is also $(0.02\%/0.06\%) = 0.3$. Therefore, if the intrinsic Cr anomalies inferred for the ordinary inclusions are also to be explained by the admixing of an anomalous component with the normal Solar System matter then a single component would be sufficient for both EK1-4-1 and ordinary inclusions.

It is important to consider whether the ^{53}Mn , ^{54}Cr and the inferred ^{53}Cr were co-produced and represent a single 'anomalous' component. We can estimate the composition of this component by assuming that it lacks ^{50}Cr and ^{52}Cr (ref. 2). Using the data on inclusion BR1 (^{54}Cr excess = 0.048% (ref. 2), $^{53}\text{Mn}/^{55}\text{Mn} = 37$ p.p.m. (ref. 2), $\text{Mn}/\text{Cr} = 0.21$ (ref. 2), and ^{53}Cr excess = 0.016% , inferred above) we arrive at the composition relative to ^{54}Cr given in the first column in Table 1. The desired composition seems to consist of comparable numbers of ^{54}Cr , ^{53}Cr and ^{53}Mn atoms but no ^{50}Cr and ^{52}Cr . It would be interesting to explore if such a pattern can be produced in a single nucleosynthetic source. As a first step, Table 1 shows a comparison of this composition with the production patterns given by Hainebach *et al.*¹² as well as the normal. It may indeed be possible to produce the desired composition in a single e-process zone characterized by η , the neutron excess parameter, somewhere between 0.071 and 0.0837; however, overproduction of ^{52}Cr could be a problem. The production pattern is too sensitive to η to allow a meaningful interpolation between the two η values where the nucleosynthetic calculations were performed. Because ^{53}Mn is neutron-poor and ^{53}Cr and ^{54}Cr are neutron-rich, one might attempt to assign these three nuclides to at least two different zones. However, the observed ^{53}Mn abundance is small, so a single-zone model remains a possibility. This particular e-process model is shown here merely to illustrate the usefulness of such an estimated composition and the possibility of producing it in a single zone. Such a zone would produce Ti highly enriched in ^{50}Ti but more ^{46}Ca than ^{48}Ca . Thus it cannot

fully account for the rough qualitative correlation observed in Allende inclusions between ^{54}Cr , ^{50}Ti and ^{48}Ca . Also, stellar nucleosynthesis considerations suggest that e-process products ejected from a massive star are expected to be a mixture of materials from zones covering a broad spectrum of η ¹³.

We have demonstrated that the offset of the Solar System point from the ^{53}Mn isochron reported in ref. 2 disagrees with the prediction of a model in which Allende inclusions formed directly from the solar nebular with homogeneous initial $^{53}\text{Mn}/^{55}\text{Mn}$ and $^{53}\text{Cr}/^{52}\text{Cr}$ ratios. This disagreement could mean that the inclusions have a more complicated history. If they have formed from an intermediate source reservoir which in turn formed from the solar reservoir, then this intermediate source must have a higher Mn/Cr than the solar ratio. Alternatively, the offset could be explained by heterogeneities in the distribution of ^{53}Mn and the inclusions had an initial ^{53}Mn twice the average Solar System value. The third possibility is that when they were formed the inclusions had an intrinsic ^{53}Cr excess of 0.02% relative to normal Solar System matter. A single component of anomalous Cr with a $^{54}\text{Cr}/^{53}\text{Cr}$ three times the normal ratio can explain not only this inferred ^{53}Cr excess and observed ^{54}Cr effects in ordinary inclusions but also the much larger Cr anomalies found in the unusual FUN inclusion EK1-4-1. If ^{53}Mn is also co-produced with this component then the nucleosynthesis of these iron-group nuclides would be much better constrained, and a single e-process zone would be a possible source.

I thank Drs D. A. Papanastassiou (Caltech) and J. L. Birk (UCSD and Univ. Paris) for comments that enabled me to improve the paper; Dr K. K. Liu of the Academia Sinica for useful discussions; and the National Science Council (Taiwan, R.O.C.) for travel grants to meetings where the ideas in this paper were first formed.

Received 19 March; accepted 19 August 1986.

1. Wasserburg, G. J. *Proto-stars and Planets II* (ed. Black, D. C. & Matthews, M. S.) 703-707 (Univ. Arizona Press, 1985).
2. Birk, J. L. & Allègre, C. J. *Geophys. Res. Lett.* **12**, 745-748 (1985).
3. Lee, T., Papanastassiou, D. A. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 109-112 (1976).
4. Nitoh, O., Kamata, M. & Honda, M. *Abstr. Vol. 4th Int. Conf. Geochron.*, Snowmass, Colorado, 308-310 (1978).
5. Lee, T. & Tera, F. *Abstr. Vol. 5th Int. Conf. Geochron.* Nikko, Japan, 212-213 (1982).
6. Lee, T. & Tera, F. *Geochim. cosmochim. Acta* **50**, 199-206 (1986).
7. Birk, J. L. & Allègre, C. J. *Geophys. Res. Lett.* **11**, 943-946 (1984).
8. Papanastassiou, D. A., Zyskind, J. L., Ngo, H. H. & Wasserburg, G. J. *Meteoritics* **20** (in the press).
9. Papanastassiou, D. A. *Astrophys. J. Lett.* (in the press).
10. Anders, E. & Ebihara, M. *Geochim. cosmochim. Acta* **46**, 2363-2380 (1982).
11. Grevesse, N. in *Frontiers of Astronomy and Astrophysics* (ed. Pallavicini, R.) 71-82 (Ital. Ast. Soc., Florence, 1984).
12. Hainebach, K. L., Clayton, D. D., Arnett, W. D. & Woosley, S. E. *Astrophys. J.* **193**, 157-168 (1974).
13. Hartmann, D., Woosley, S. E. & El Eid, M. F. *Astrophys. J.* **297**, 837 (1985).

EISCAT observations of the E-region semidiurnal tide

T. S. Virdi, G. O. L. Jones & P. J. S. Williams

University College of Wales, Department of Physics, Penglais, Aberystwyth SY23 3BZ, UK

Many observers have reported a semidiurnal tide in the E-region of the ionosphere^{1,7}. Theory demonstrates the possibility of several tidal modes with the relative strength of each mode depending on the efficiency of generation in the stratosphere, the attenuation suffered in propagating through the mesosphere and mode-coupling in regions of nonlinearity. To determine which modes actually occur it is important to observe the tide at all heights in the E-region and at a range of latitudes. EISCAT, the European incoherent scatter radar^{2,3}, is an ideal facility to make such observations at 69.5° N. At this latitude, measurements are usually

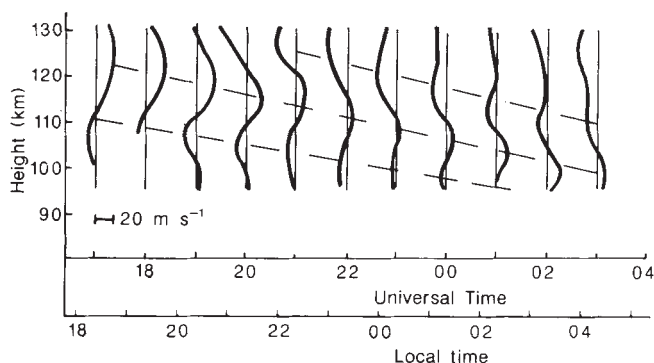


Fig. 1 Height profiles of ion velocity parallel to the magnetic field line, 14-15 February 1985.

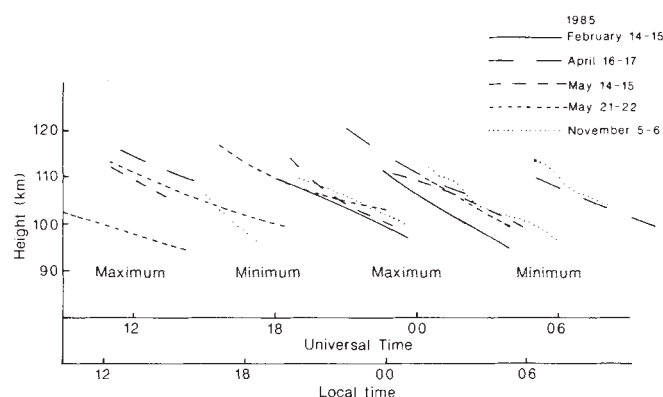


Fig. 2 The descending phase-fronts of the semidiurnal tide as demonstrated by the maxima (upward) and minima (downward) of ion velocity parallel to the magnetic field above Tromsø on five days during 1985.

perturbed by strong auroral heating, but near sunspot minimum auroral disturbances are sometimes very small and the underlying pattern of the semidiurnal tides can be observed. Such conditions occurred in 1985 and here we report observations on five different days that illustrated the dominance of the (2,4) mode at high latitudes.

At altitudes below 120 km in the ionosphere, the ion-neutral collision frequency is much greater than the ion-gyro frequency so that movement of the ion population is mainly controlled by the neutral atmosphere. Measurements of ion velocity can therefore monitor neutral winds.

The data presented here were taken while EISCAT was performing the Common Programme CP1. The antenna at Tromsø transmitted along the magnetic field line and observed the whole E-region with a height resolution of 3 km. Remote antennas at Kiruna and Sodankylä observed the signals scattered from heights of 101, 110, 119, 131 and 312 km. The data were analysed to provide height profiles of electron density, electron and ion temperature and plasma velocity.

Figure 1 shows hourly averages of plasma velocity along the field line, $v_{\parallel}$, as a function of height during 14-15 February. A sinusoidal variation with height is observed, and this pattern descends steadily at a rate of 2-3 km h⁻¹.

In Fig. 2 the heights of the maxima and minima in $v_{\parallel}$ are plotted against time for each of the five days. The pattern is consistent, and small differences from day to day can be accounted for by seasonal changes in the neutral atmosphere, day-to-day variations in the relative strength of the different diurnal and semidiurnal modes, the effects of auroral electric fields, especially above 120 km, and random errors in the measurements themselves.

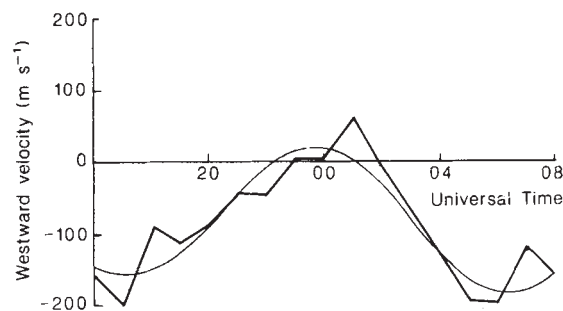
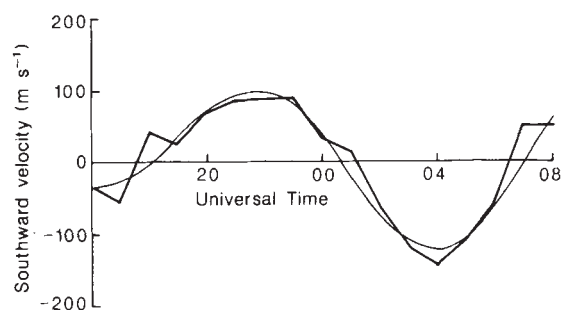


Fig. 3 Horizontal velocity in the southward and westward direction at a height of 101 km, 5-6 November 1984. Heavy line, observed velocity; tight line, best-fitting diurnal and semidiurnal tide.

To determine the actual wind velocity associated with the tidal modes we calculated the full vector of plasma velocity, $\mathbf{v}_p$, by combining the separate components of velocity measured at the three EISCAT stations and applying a small correction for the electric field, derived from measurements at 312 km. Figure 3 shows typical results: the horizontal north-south and east-west velocities on 5-6 November are plotted against time for a height of 101 km, showing the semidiurnal tide.

Theory^{4,5} predicts that the tidal mode (m,n) shows a downward phase propagation with a vertical wavelength given by

$$\lambda_{z,mn} = 2\pi H / \sqrt{\frac{H}{h_{mn}} \left(\frac{\gamma-1}{\gamma} + \frac{dH}{dz} \right) - \frac{1}{4}}$$

where H is the scale height of the atmosphere at height z , γ is the ratio of specific heats at constant pressure and constant volume, and h_{mn} is the equivalent depth of the tidal mode. In Table 1 average values of h_{mn} and $\lambda_{z,mn}$ are quoted for various semidiurnal modes at a height of 110 km, corresponding to the MSIS83 model of the atmosphere⁶ for the five days.

In Fig. 4 the value of λ_z is used to predict the rate of descent of the phase front for each mode. The average rate of descent of the phase front actually observed for $v_{\parallel}$ is also plotted.

The pattern descends from 120 to 100 km at an average rate of 2.5 km h⁻¹ suggesting a vertical wavelength of about 30 km. We can therefore eliminate the (2,2) and (2,3) modes, but allowing for uncertainties in the atmospheric model, especially

Table 1 Predicted vertical wavelength

Mode (m,n)	h_{mn} (km)	$\langle \lambda_{z,mn} \rangle$ (km)
(2,2)	7.85	96
(2,3)	3.67	53
(2,4)	2.11	38
(2,5)	1.37	30
(2,6)	0.96	25

Table 2 Observed vertical wavelength

Date	Time of maximum (UT)					
	North-south			East-west		
	101 km	110 km	119 km	101 km	110 km	119 km
16 April	—	6.7	—	—	9.0	6.5
14 May	—	6.4	4.7	10.9	8.8	4.6
21 May	8.5	6.3	—	11.8	8.8	6.2
5 November	10.2	6.8	—	11.9	9.6	—
Average	9.4	6.6	4.7	11.5	9.0	5.8
Wavelength (km)	44.6			37.5		

in the estimate of dH/dz , and for the perturbing effects of electric fields the observations are compatible with (2,4), (2,5) or (2,6).

More precise evidence involves measurements of the full vector of ion velocity whenever tristatic data are available. In the E-region v_p is largely determined by (1) the semidiurnal tide in the neutral atmosphere, (2) the diurnal tide, and (3) the effect of ionospheric electric fields. At 101 km the effects of the electric field are very small and the amplitude and phase of the semidiurnal and diurnal modes can be determined by a least-squares fit. At greater heights, as the ion-neutral collision frequency decreases, the ion velocity is increasingly affected by the electric field. At 110 and 119 km, however, it is possible to make an accurate correction for this and the phase of the semidiurnal tide can be determined. Table 2 summarizes results for all days with adequate data, and Fig. 4 shows the average time of minimum southward component at each height.

The variation of phase with height corresponds to a vertical

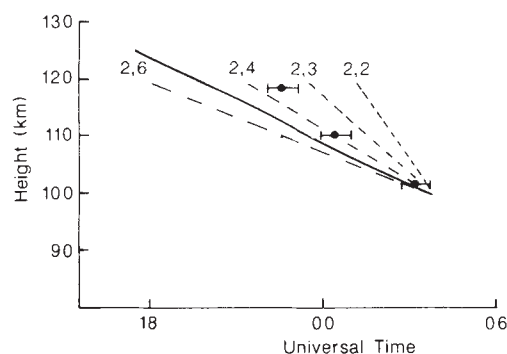


Fig. 4 The predicted rate of descent of several semidiurnal modes (broken lines) and the observed rate-of-descent of the minima of the ion velocity parallel to the magnetic field line (continuous line) and the time of minimum southward horizontal velocity at 101, 110 and 119 km (points with error bars).

wavelength of 41 ± 2 km, which strongly suggests the dominance of the (2,4) mode at high latitudes.

We thank the Director and staff of EISCAT for making the observations and analysing the data. EISCAT is supported by the research councils of France, Finland, Germany, Norway, Sweden and the UK.

Received 4 July; accepted 1 September 1986.

1. Alcayde, D. & Bernard, R. *J. Atmos. terr. Phys.* **44**, 95-109 (1982).
2. Folkestad, K., Hagfors, T. & Westerlund, S. *Radio Sci.* **18**, 867-879 (1983).
3. Rishbeth, H. & Williams, P. J. S. Q. *J. R. astr. Soc.* **26**, 478-512 (1985).
4. Wilkes, M. V. *Oscillations of the Earth's Atmosphere* (Cambridge University Press, 1949).
5. Chapman, S. & Lindzen, R. S. *Atmospheric Tides* (Reidel, Dordrecht, 1970).
6. Hedin, A. E. *J. geophys. Res.* **88**, 10,170-10,188 (1983).
7. Kirkwood, S. C. *J. Atmos. terr. Phys.* **48**, 817-826 (1986).

Long-range transport and deposition of acidic nitrogen species in north-west Europe

Richard G. Derwent* & Katrin Nodop†

* Modelling and Assessments Group, Environmental and Medical Sciences Division, Harwell Laboratory, Oxfordshire OX11 0RA, UK
† EMEP Chemical Coordinating Centre, Norwegian Institute for Air Research, Lillestrøm, Norway

The mechanisms of the long-range transport and deposition of acidic nitrogen compounds over north-west Europe are not yet well understood. Aerosol nitrate concentrations have been steadily rising in southern England¹ over the past three decades, as have the nitrate concentrations in precipitation at almost all sites in north-west Europe, including the most remote^{2,3}. The life cycle of acidic nitrogen compounds potentially involves a wide range of trace constituents including nitric oxide (NO), nitrogen dioxide (NO₂), gaseous nitric acid (HNO₃), higher nitrogen oxides (such as N₂O₅), nitrate aerosol and nitrate in precipitation. Here we show that a simple model, incorporating an assumed life cycle of nitrogen-containing trace gases and the known NO sources, can account for much of the nitrate in rain observed over north-west Europe.

The spacial distributions of dry and wet deposition of acidic nitrogen compounds were calculated using a simple trajectory model, the details of which are described elsewhere⁴. The formulation of the model is based on the EMEP (European Monitoring and Evaluation Programme) long-range sulphur transport model⁵, although we have made some major simplifications. These have involved the assumption of straight line trajectories; mixing heights held constant at 1,000 m; windspeeds at 8 m s⁻¹;

and statistically averaged precipitation-removal rates constant throughout Europe. Figure 1 shows the life cycle of the acidic oxidized nitrogen compounds that has been used and the rate coefficients used to link each of the modelled species. The model consists of a set of coupled differential equations, one for each species including ozone (O₃), representing mass conservation and was solved using FACSIMILE, a variable-order Gear's method program⁶ on an IBM PC XT microcomputer.

The major source of acidic nitrogen compounds in north-west Europe is the emission of NO from motor vehicles and stationary combustion. The calculations described here used the NO emissions⁷ on the 150 km × 150 km EMEP grid for 1978 from

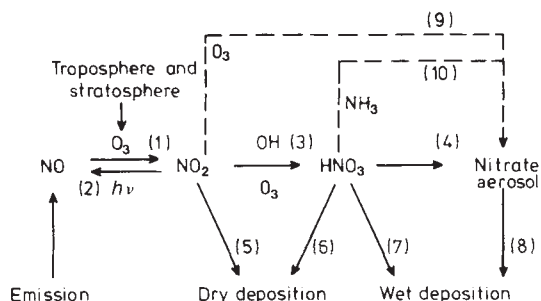


Fig. 1 The life cycle of the nitrogen-containing trace gases and the rate coefficients for the interconversion processes between them. (1) 3.8×10^{-4} p.p.b.⁻¹ s⁻¹; (2) 2.1×10^{-3} s⁻¹; (3) 3.4×10^{-5} s⁻¹; (4) 3.0×10^{-5} s⁻¹; (5) 4.0×10^{-6} s⁻¹; (6) 2.0×10^{-5} s⁻¹; (7) 9.0×10^{-6} s⁻¹; (8) 1.3×10^{-5} s⁻¹; (9) 1.0×10^{-6} p.p.b.⁻¹ s⁻¹; (10) 2.4×10^{-4} p.p.b.⁻¹ s⁻¹. The solid lines refer to processes in the base model and the dotted lines to those in the base model + NH₃ + night-time chemistry.



Fig. 2 The calculated nitrate concentration in rain over north-west Europe in $\text{mg N l}^{-1} \text{NO}_3$ in: *a*, the base model; *b*, the base model + NH_3 chemistry; *c*, the base model + NH_3 + night-time chemistry.

stationary combustion alone. The contribution from motor vehicles to total NO emissions has yet to be specifically treated on the EMEP grid but it is known to be significant overall and could therefore influence both the magnitude and spatial distribution of the emissions. The model allows for the oxidation of NO to NO_2 by the well known ozone reaction and then to gaseous nitric acid by reaction with OH radicals. A tropospheric background ozone concentration of 30 p.p.b. (30 parts per 10^9 , $60 \mu\text{g m}^{-3}$) has been employed at the start of each trajectory and is replenished by vertical exchange along the trajectory. The OH radical concentration was taken to be that same concentration ($2 \times 10^6 \text{ molecule cm}^{-3}$) which would account for the SO_2 oxidation in the EMEP sulphur transport model⁵. In the assumed life cycle, HNO_3 is subsequently scavenged by alkaline soil particles of the coarse atmospheric aerosol and by sea-salt aerosol to form nitrate aerosol. Dry deposition arises through removal of NO_2 and HNO_3 at the Earth's surface and wet deposition through wash-out of HNO_3 and rain-out of nitrate aerosol. Wet deposition was estimated using a constant scavenging coefficient⁴ (Fig. 1) and an annual rainfall of 1,000 mm. None of the processes in Fig. 1 is yet adequately defined on a European scale, including the emissions, for oxidized nitrogen species.

The concentrations of NO, NO_2 , HNO_3 and nitrate aerosol together with the dry and wet deposition rates of the oxidized nitrogen compounds at a given location were obtained by solving the differential equations describing their production and loss for 24 separate 96-h trajectories through to that location, three trajectories within each of eight wind direction sectors 45° wide. The long-term mean behaviour at this location was obtained by weighting the results for each trajectory by the frequency of the 96-h back trajectories assigned to each 45° sector in the daily analysis completed by the Meteorological Synthesizing Centre-W for the 636-day period from 1 February 1981 to 29 October 1982 (ref. 8). This process was repeated for each of the 64 UN ECE EMEP receptor sites located within the study region for which the above daily sector analysis was available.

The calculated concentrations of nitrate in precipitation at each of the 64 receptor sites were plotted out and contours were drawn, joining places of identical rainfall composition (see Fig. 2*a*). Most of north-west Europe lay within the $0.1 \text{ mg N l}^{-1} \text{NO}_3$ contour under the base model assumptions. The total area integrated dry and wet deposition of oxidized nitrogen compounds to the $2 \times 10^6 \text{ km}^2$ land area of Belgium, Denmark,

France, FRG, Ireland, Netherlands, Norway, Sweden and the UK amounted to 420 and $450 \times 10^9 \text{ g N yr}^{-1}$. The mean residence time for oxidized nitrogen in this base model was found to be 1.2 days, corresponding to a mean transport distance of 800 km. The calculated nitrate concentrations in precipitation in Fig. 2*a* showed a systematic underestimation of the observed⁹ concentration data in Fig. 3.

In formulating the base model, two additional processes have been specifically omitted: (1) scavenging of gaseous nitric acid by ammonia and its subsequent condensation to form ammonium nitrate aerosol; (2) reaction of NO_2 with ozone at night-time to form N_2O_5 and its subsequent reaction with water droplets to form nitrate aerosol. We now illustrate the potential importance of these two processes on the spatial distribution of wet nitrate deposition.

The formation of ammonium nitrate aerosol was treated by including an additional differential equation for ammonia, representing its emission from animal and industrial sources on the same $150 \times 150 \text{ km}$ grid¹⁰, its dry deposition and reaction with HNO_3 to form ammonium nitrate aerosol and its subsequent wet deposition. The modelling problem was simplified by assuming no volatilization and redissociation of ammonium nitrate aerosol and by neglecting any competitive scavenging of NH_3 by sulphuric or hydrochloric acid gases. These assumptions should exaggerate the importance of ammonium nitrate aerosol formation on the nitrogen life cycle. The rate coefficients for the various processes were estimated as few of the relevant processes have yet been studied in any detail except for the thermodynamics of ammonium nitrate formation¹¹.

The formation of N_2O_5 at night-time was determined using the known rate coefficient of the $\text{NO}_2 + \text{O}_3$ reaction¹². The NO_3 produced was assumed to be scavenged to form N_2O_5 which, in turn, reacted heterogeneously with water droplets to form nitrate aerosol. Constant night-time conditions were taken so that our assumptions should exaggerate the importance of this route in parallel with the optimistic assumptions also made for the ammonium nitrate aerosol formation route. The assumptions made for the two additional routes, distinguished from the base model in Fig. 1, should therefore be seen to be optimistic but not unreasonably so.

Figure 2*b,c* shows the nitrate concentrations in rain calculated for the base model + NH_3 and the base model + NH_3 + night-time chemistry. A slight widening of the area covered by the $0.1 \text{ mg N l}^{-1} \text{NO}_3$ contour is apparent compared with the base model

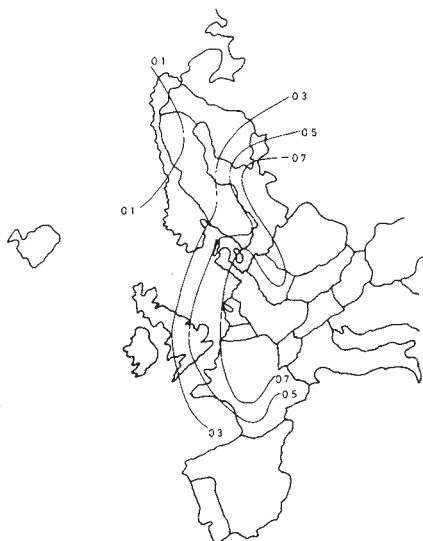


Fig. 3 The observed⁸ nitrate concentration in rain over north-west Europe in $\text{mg N l}^{-1} \text{NO}_3$.

results in Fig. 2a. However, the main change is the marked increase in the magnitude of the peak nitrate concentration in precipitation from 0.64 to 0.82 and $1.08 \text{ mg N l}^{-1} \text{NO}_3$. The location of the peak nitrate concentration with precipitation remains in the centre of the FRG with a band of high concentration extending towards the UK through the low countries. A marked increase in the calculated nitrate content of precipitation is noted throughout all the region studied, completely removing the previously noted systematic underestimation of the observations in Fig. 3. All these remarks concern the wet deposition of oxidized nitrogen compounds and not the total wet nitrogen deposition which would include a further contribution specifically from the deposition of ammonia and its derivatives.

The effect of the ammonia and night-time chemistry has been to reduce the total area integrated dry deposition of oxidized nitrogen compounds to $140 \times 10^9 \text{ g N yr}^{-1}$ and increase their wet deposition to $710 \times 10^9 \text{ g N yr}^{-1}$. The total wet and dry deposition has, however, remained unchanged and all that has been affected is a change in the deposition route from dry deposition to wet deposition. The average residence time of oxidized nitrogen has increased to 1.3 days. Total wet and dry deposition of oxidized nitrogen compounds is controlled by the NO emission strength which has been left unchanged, but the detailed chemistry controls the split between wet and dry deposition. This split is crucial because only wet deposition is directly measurable over north-west Europe. Furthermore, wet and dry deposition may have different ecological effects.

The model results were compared with measurements at 25 EMEP sites for which nitrate in precipitation data are available for either 1981 or 1982 or both. Table 1 shows how the model results correlate with the observations revealing a high degree of correlation ($R^2=0.8$) between the model calculations and observations for the 25 sites. The slopes are significantly different between the various model experiments and approach unity only when the ammonia and night-time chemistry is introduced. In all cases, the intercepts are not statistically different from zero indicating the apparent absence of background contribution to nitrate in precipitation. This is an apparent contrast with the analogous situation for sulphate in precipitation⁵.

Table 1 Comparison of the observed and model calculated values of nitrate in precipitation at 25 EMEP sites in north-west Europe

Correlations	Slope	Intercept (mg N l^{-1} as NO_3)	Correlation coefficient (R^2)
Base model on data	0.62 ± 0.1	0.02 ± 0.05	0.78
Base model + NH_3 on data	0.80 ± 0.1	0.03 ± 0.07	0.78
Base model + NH_3 + night-time chemistry on data	1.03 ± 0.2	0.01 ± 0.09	0.76

The simple model, with the assumed life cycle and the known NO sources over north-west Europe are therefore able to account for much of the nitrate in rain observed over north-west Europe. However, an adequate explanation requires a better meteorological description and a more detailed understanding of the atmospheric chemistry of NO_x , ammonia and ozone than has been used here and will require a much improved database for ammonia and ozone across north-west Europe in addition to that for NO , NO_2 , HNO_3 and nitrate aerosol. Furthermore, as European NO emissions have almost certainly increased over the past 30 years then total wet and dry deposition will have risen proportionately. However, because of possible increased NH_3 emissions¹³ and possible increased background ozone concentrations¹⁴, this work demonstrates that the split between wet and dry deposition may also have changed markedly leading to a more dramatic increase in wet nitrate deposition compared with total deposition. Future trends in wet nitrate deposition may not necessarily be dictated solely by NO_x emissions and by the application of NO_x control technologies in vehicles and large combustion plant unless steps are taken to understand and control future NH_3 and ozone concentrations and their possible changes over north-west Europe.

This work was supported in the UK by the Department of Environment and at the Norwegian Institute for Air Research through the UN ECE EMEP Programme and through a contract from the Umweldbundesamt of the FRG. We acknowledge discussions with John Garland, Frank Leeuw, Oystein Hov and Anton Eliassen.

Received 28 July; accepted 3 October 1986.

1. Salmon, L., Atkins, D. H. F., Fisher, E. M. R., Healy, C. & Law, D. V. *Sci. Total Envir.* 9, 161-200 (1978).
2. Rodhe, H. K. *Skogs-ö Lantbr Akad. Tidskr. Suppl.* 14, 32-36 (1982).
3. Kallend, A. S., Marsh, A. R. W., Pickles, J. H. & Proctor, M. V. *Atmos. Envir.* 17, 127-137 (1983).
4. Derwent, R. G. *The Nitrogen Budget for the United Kingdom and North-West Europe* (ETSU Rep. R.37, Harwell Laboratory, 1986).
5. Eliassen, A. *Atmos. Envir.* 12, 479-487 (1978).
6. Curtis, A. R. & Sweetenham, W. P. *FACSIMILE Release H User's Manual* (AERE Rep. R11771, Harwell Laboratory, 1985).
7. Semb, A. & Amble, E. *Emission of Nitrogen Oxides from Fossil Fuel Combustion in Europe* (NILU Rep. No. 13/81, Lillestrøm, Norway, 1981).
8. Lemshaus, J., Saltbones, J. & Eliassen, A. *Deposition Patterns and Transport Sector Analysis for a Four-year-period* (The Norwegian Meteorological Institute Report EMEP/MSC-W 1/85, Oslo, 1985).
9. *Long Range Transport of Air Pollutants. Monitor 1984*, 1-134 (National Swedish Environmental Protection Board, Solna, Sweden, 1984).
10. Buijsmann, E., Mass, J. F. M. & Asman, W. A. H. *Ammonia emission in Europe* (Institute for Meteorology and Oceanography, State University of Utrecht IMOU Rep. R-85-2, Netherlands, 1985).
11. Bassett, M. E. & Seinfeld, J. H. *Atmos. Envir.* 17, 2237-2252 (1983).
12. Baulch, D. L. *et al. J. Phys. Chem. Ref. Data* 13, 1259-1375 (1984).
13. ApSimon, H. M., Kruse, M. & Bell, N. *Ammonia Emissions and their Role in Acid Deposition* (Imperial College of Science and Technology, London, 1986).
14. Penkett, S. A. *Nature* 311, 14-15 (1984).

Baryonyx, a remarkable new theropod dinosaur

Alan J. Charig & Angela C. Milner

Department of Palaeontology, British Museum (Natural History), Cromwell Road, London SW7 5BD, UK

An extremely large claw bone, some 30 cm long, was found in Wealden (Lower Cretaceous) deposits in a Surrey claypit in January 1983. This led to the discovery the following month of the well-preserved skeleton of a new large theropod dinosaur. Only one other theropod specimen comprising more than a few bones had ever been found in Britain, and that discovery was more than a century ago. Indeed, no large theropod, reasonably complete, had previously been discovered in Lower Cretaceous rocks anywhere in the world. Our study so far suggests that the Surrey dinosaur was a typical large theropod in certain respects, resembling, for example *Allosaurus*¹. In several other respects, however, it differs sufficiently from all known dinosaurs to merit designation as the representative of a new species, genus and family.

More than three years have elapsed since the skeleton was excavated by the British Museum (Natural History) in May–June 1983. We estimate that 70% of the bones collected have already been prepared from the encasing siltstone concretions, although many require further treatment. The immense amount of work involved suggests that it will be at least another four years before a full account of this important creature can be published; we have therefore decided to publish a preliminary account immediately, bestowing a name on the animal.

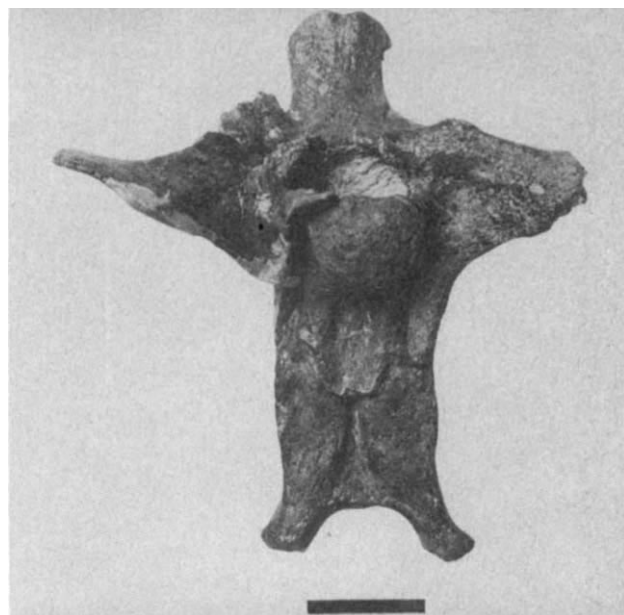


Fig. 2 *Baryonyx walkeri* holotype, no. R. 9951: Occiput, from behind. Scale bar, 5 cm.

Family *Baryonychidae* nov.

Diagnosis. As for genus *Baryonyx*, below.

Genus *Baryonyx* nov.

Derivation of name. Gk, βαρύ, heavy, strong; ὄνυξ, talon, claw.

Type-species. *B. walkeri* nov.

Diagnosis. As for species *B. walkeri*, below.

Species *Baryonyx walkeri* nov.

Derivation of name. For Mr William J. Walker, who found the first indication of the skeleton, namely the large claw bone (ungual phalanx).

Holotype. B.M.(N.H.) Palaeo. Dept. R. 9951.

Material. Conjoined premaxillae, front of left maxilla, conjoined nasals, lacrimal and parts of adjacent bones, frontals, anterior end of braincase, occiput, left dentary, some post-dentary bones of lower jaw; teeth in both jaws and in isolation; axis, most of the other cervical vertebrae, some dorsal vertebrae, one caudal vertebra; cervical rib, dorsal ribs, gastralia, chevrons; both scapulae, both coracoids, possible clavicle; both humeri, phalanges of manus including unguals (one of them, not certainly from manus, extremely large); pieces of ilia and pubes, ischium; proximal end of left femur and distal end of right, left fibula lacking distal end; right calcaneum; distal ends of metatarsals, phalanges of pes including unguals; various unidentified pieces; suspected gastroliths.

Locality. Claypit at Smokejacks Brickworks (Ockley Brick Company Limited), Wallis Wood, Ockley, near Dorking, Surrey, England. Map reference TQ13: 113 373.

Horizon. Weald Clay, near the base of the Barremian, Wealden (Lower Cretaceous); approximate age 124 million years.

The bones, though largely disarticulated and scattered, were all found within an area 5 m × 2 m; except where previously disturbed by the bulldozer, they were generally not far from their natural positions, that is most of the pieces of the skull, pectoral girdle and fore-limbs were at one end and most of the pelvic girdle and hind-limbs at the other. They were not crushed or distorted to any significant degree. Many had been broken cleanly before fossilization; most had become encased in sideritic siltstone nodules before the broken ends had weathered or eroded, but a few lay unprotected in clay. So far we have recognized few other vertebrate remains mixed with those of this one individual of *Baryonyx*; these include fish teeth and scales, the humerus of a small *Iguanodon* and a small phalanx. The hardness of the siltstone concretions necessitated the com-

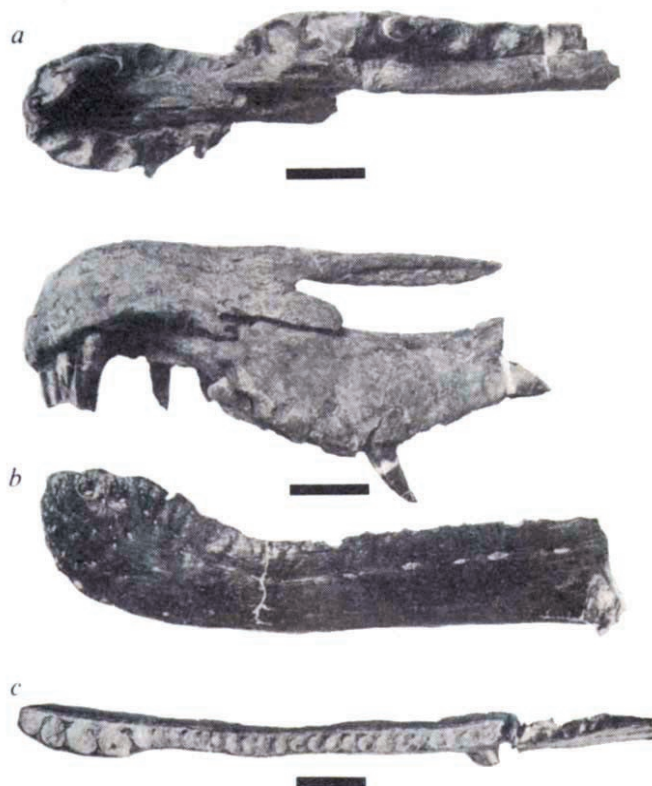


Fig. 1 *Baryonyx walkeri* holotype, no. R. 9951: a, Conjoined premaxillae together with left maxilla (all incomplete in articulation; in occlusal (ventral) view. b, Premaxillae and left maxilla as above, also with incomplete left dentary in approximately correct relative orientation; from the left side. c, Left dentary; in occlusal (dorsal) view, to show the full complement of tooth sockets. Scale bars, 5 cm.

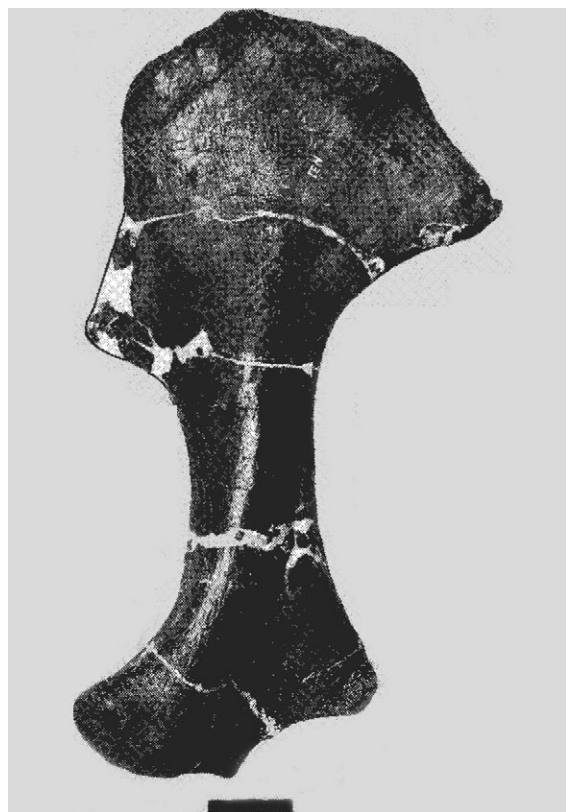


Fig. 3 *Baryonyx walkeri* holotype, no. R.9951: Left humerus, in anterolateral view. Scale bar, 5 cm.

bined use of mechanical (Vacublast, air-abrasive, diamond saws and dental mallets) with chemical (thioglycolic acid²) methods of preparation.

Diagnosis. The new theropod is distinguished by a unique combination of characters. These include: the prenasal extension of the snout into an extremely narrow rostrum with a spatulate, horizontal expansion at its end, slightly downturned in lateral view (Fig. 1); the long, low external naris, a flat triangle with its apex directed forwards, situated far back from the front of the snout and ascending gently towards the rear; the seemingly mobile articulation of the premaxilla and maxilla above a sub-rostral notch (Fig. 1); the small median knob at the posterior end of the conjoined nasals, cruciform in dorsal view, with the anterior limb of the cross drawn forwards into a low thin median crest; the deep occiput (Fig. 2) with paroccipital processes directed horizontally outwards and with basiptyergoid processes descending far below the basioccipital and diverging laterally only slightly; the unusually high number of teeth

$$\frac{pm\ 7 + m?}{d\ 32}$$

contrast with the usual theropod count of

$$\frac{pm\ 5 + m\ 16}{d\ 16}$$

the largest being the third in the premaxilla and the third and fourth in the dentary, the crowns lightly fluted on the lingual side, with very fine serrations (about 7 per millimetre) on the anterior and posterior keels, the roots long and slender; no obvious interdental plates (unlike *Megalosaurus*, where they are very conspicuous, high and separate from each other); the long, strongly opisthocoelous cervical vertebrae with short neural spines, with well developed epipophyses, and with ends of the centra not 'offset' so that when they are correctly apposed the neck does not show the typical theropod upward curve; the



Fig. 4 *Baryonyx walkeri* holotype, no. R.9951: a, 'Normal' digit, presumably of hand. b, Especially large ungual phalanx. Length of scale bar 5 cm.

relatively well-developed humerus (Fig. 3), with both ends broadly expanded but extremely flattened, offset 35° against each other, shaft massive and almost straight; the ilium with a long, straight posterior process; and at least one pair of huge talons (probably on the hands, Fig. 4). There are other distinguishing characters of a more esoteric nature.

Baryonyx shows a strange mixture of primitive and specialized characters. The downturned tip to the snout, the apparent mobility between premaxilla and maxilla and the subrostral notch, the elongation of the cervical vertebrae and the form of the posterior process to the ilium are typical of many thecodontians; the humerus is also primitive in some respects. It is remarkable that so many characters frequently found in Triassic archosaurs should still be present in those from the Early Cretaceous. In contrast, the opisthocoely of the cervical vertebrae and their epipophyses, the lack of a supinator process and ectepicondylar groove to the humerus, the inturned head to the femur and the apparent absence of dermal armour clearly demonstrate that *Baryonyx* is a theropod (as currently defined) with the 'fully improved' stance and gait that distinguish dinosaurs from their thecodontian ancestors. It is not a crocodilian, for neither the skull, the cervical centra, the pelvis nor the ankle shows any of the appropriate specializations. Even so, if the animal is to be considered a theropod, it possesses unique specializations of its own: the terminal 'rosette' to the upper jaw, the extreme narrowness of the skull just behind it, the median knob and crest on the nasals, the large number of teeth (double that of a typical theropod), the lack of an upward curvature to the neck, and the huge offensive/defensive talon. It might also have been quadrupedal to some extent (see below), in which case such stance and gait—whether primitive or secondarily derived—would certainly be unique among theropods.

Therefore we propose that the most proper systematic position for *Baryonyx*, on the evidence presently available, is in a new family (Baryonychidae) of the Theropoda; the information currently available does not indicate a particular relationship to any other group of theropods. (The nature of the articulation between premaxilla and maxilla is somewhat reminiscent of *Dilophosaurus*^{3,4} from the Lower Jurassic Kayenta Formation of Arizona, placed by Welles⁴ in the Hapticosauridae, but *Dilophosaurus* has paired crests on top of its head, whereas *Baryonyx* appears to have only one.)

The only other material that may be assigned to the same family are two fragmentary snouts from the Aptian of Niger, described⁵ as the mandibular symphyses of a spinosaurid. They are, however, almost identical to the conjoined premaxillae of

Baryonyx, and comparison with *Spinosaurus*⁶ does not suggest spinosaurid affinities.

The possible life habits of *Baryonyx* are unclear. The elongated skull, the long neck and the form and strong development of the humerus might all suggest quadrupedality, facultative at least, even though the flattened slender ends of the humerus makes it unlikely that the stance and gait were predominantly quadrupedal. The lateral position of the external nares, far back from the tip of the snout, seems to preclude an aquatic or amphibious existence, and the form of the postcranial skeleton is wholly suggestive of life on land. Nevertheless the elongated snout with its terminal expansion and the large number of finely serrated teeth may well be indicative of ichthyophagy. Taquet's article⁶ on the Niger theropod made the same suggestion, comparing the elongated snouts with those of the Recent *Gavialis gangeticus* and commenting that "il est tentant d'imaginer des Dinosaures Spinosauridés pêchant le long des fleuves ou au bord des lacs à la manière des Hérons ou des Cigognes." We independently conceived a similar idea for *Baryonyx*, though as a quadrupedal predator crouching on the bank rather than as a biped stalking through the shallows. A scavenging habit has also been suggested for the new dinosaur, but the evidence for either mode of life is inconclusive. The associated fossil fauna is dominated by the remains of the herbivorous ornithomimid dinosaur *Iguanodon* and of many insects⁷, the latter under continuing study by Jarzembowski.

Whether the huge talon was one of an only pair remains unknown; likewise whether it was borne by the fore-limb or by the hind. The ungual is exactly like those of *Allosaurus*, only much bigger. Clearly it was a weapon of offence and/or defence; but it seems likely that the great size of *Baryonyx* (it may have

weighed a couple of tons or more) would have precluded any possibility of emulating its contemporary *Deinonychus* from the Cloverly Formation of Montana⁸, a biped of much smaller dimensions which bore a large, laterally-compressed, sickle-like claw on the second digit of the pes and which balanced on the other hind-leg while slashing at its victim. If, however, the great talon of *Baryonyx* was on the manus, it could be connected with the powerful development of the humerus. All this must have some bearing on facultative quadrupedality (the claw would have had to be held clear of the ground while the animal was walking) and on ichthyophagy; could *Baryonyx* have used the claw, like a grizzly bear, for 'gaffing' fish out of the water? *Baryonyx* is the dinosaur that in 1983 received world-wide publicity, and became nicknamed 'Claws'. We are expecting its remains to provide a great deal more information over the next few years.

Our thanks are due to the many people who helped in different ways, too numerous to specify. They include Mr William Walker, the Directors and staff of the Ockley Brick Company Limited, the collecting teams (Museum staff and others), the Museum photographers, Mr Philip Palmer, and Mr Ronald Croucher and his colleagues in the Museum's Palaeontological Laboratory.

Received 4 August; accepted 1 October 1986.

1. Madsen, J. H. Jr *Bull. Utah geol. miner. Surv.* **109**, i-xii, 1-163 (1976).
2. Howie, F. M. P. *Curator* **17**, 159-166 (1974).
3. Welles, S. P. *Bull. geol. Soc. Am.* **65**, 591-598 (1954).
4. Welles, S. P. *Palaeontographica*, Abt. A **185**, 85-180 (1984).
5. Taquet, P. C. r. heb. Séanc. Acad. Sci., Paris **299**, 217-222 (1984).
6. Stromer, E. *Abh. Bayer. Akad. Wiss.* **28**, 1-32 (1915).
7. Jarzembowski, E. A. *Proc. Geol. Ass.* **87**, 443-446 (1977).
8. Ostrom, J. H. *Bull. Peabody Mus. nat. Hist.* **30**, 1-165 (1969).

Functional architecture of cortex revealed by optical imaging of intrinsic signals

Amiram Grinvald*†‡, Edmund Lieke*, Ron D. Frostig*, Charles D. Gilbert* & Torsten N. Wiesel*

* Laboratory of Neurobiology, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

† Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel

‡ T. J. Watson Research Center, IBM, Yorktown Heights, New York 10598, USA

Optical imaging of cortical activity offers several advantages over conventional electrophysiological and anatomical techniques. One can map a relatively large region, obtain successive maps to different stimuli in the same cortical area and follow variations in response over time. In the intact mammalian brain this imaging has been accomplished with the aid of voltage sensitive dyes¹⁻⁵. However, it has been known for many years that some intrinsic changes in the optical properties of the tissue are dependent on electrical or metabolic activity⁶⁻¹³. Here we show that these changes can be used to study the functional architecture of cortex. Optical maps of whisker barrels in the rat and the orientation columns in the cat visual cortex, obtained by reflection measurements of the intrinsic signal, were confirmed with voltage sensitive dyes or by electrophysiological recordings. In addition, we describe an intrinsic signal originating from small arteries which can be used to investigate the communication between local neuronal activity and the microvasculature. One advantage of the method is that it is non-invasive and does not require dyes, a clear benefit for clinical applications.

The imaging of changes in membrane potential in central nervous system (CNS) preparations with voltage sensitive dyes,

which are usually associated with fast signals, is often distorted by much slower intrinsic signals^{6,11}. The observation of only slow signals from the surface of monkey striate cortex stained with a voltage sensitive dye¹⁴ prompted us to test whether slow reflectance signals could be observed without using any dye, and to compare the relative merits of intrinsic and dye-related imaging. For optical recording from the rat somatosensory cortex, the exposed cortex was viewed with a compound microscope equipped with fluorescence epi-illumination through a glass-coverslip sealed chamber. The rat somatosensory cortex includes a region that is subdivided into smaller areas, known as barrels, each of which responds to stimulation of a single mystacial whisker¹⁵. Using the photodetector array¹⁶ to monitor changes in reflected light from an active barrel field, we observed a decrease in reflectance in response to whisker stimulation. In dye stained preparations, this intrinsic signal was distinguished from the dye signal by measurement at a wavelength range outside the absorption band of the dye⁶. This made it possible to compare directly maps obtained by the reflected light and voltage sensitive dye signals in response to stimulation of one whisker. Both showed signals in a well-localized cortical region and the centres of the active areas overlapped very closely (Fig. 1a, b). Although the reflected light signal was much slower than the dye signal (Fig. 2a), representing different processes, it was still as useful for visualizing the active cortical region.

To determine the optimal conditions for the measurement and to characterize the signal source, we examined the time course and wavelength dependence of the signal. The reflected light signal in the rat somatosensory and in the cat and monkey visual cortex had a slower rise time (>1 s) than the dye related signals (Fig. 2a, b). This time course is clearly very different from neuronal electrical activity. By comparing the optical signals in response to long and short stimuli we found that the rise time and time to peak were nearly independent of stimulus duration (Fig. 2d). Examination of the relaxation time course of the signal was important for choosing the appropriate inter-stimulus interval. Short duration stimuli (0.5-2 seconds) pro-

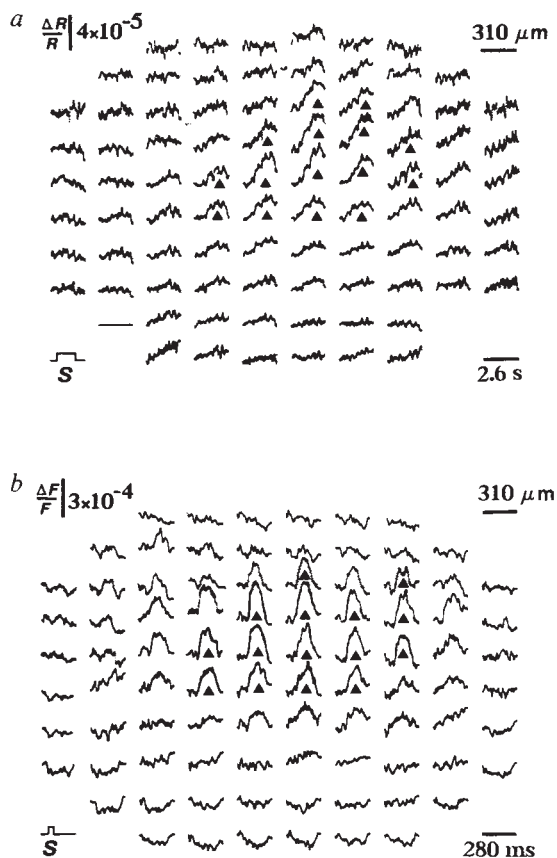


Fig. 1 Mapping of a single barrel in rat cortex with the intrinsic signal and its confirmation with a voltage-sensitive dye. *a*, The spatial distribution of the reflected light signals in response to a mechanical stimulation of a single whisker. Simultaneous intensity changes in reflected light were measured with 96 photodetectors, averaged over 200 trials. Signals marked by triangles are the largest 15 signals selected by computer integration. A wavelength of 665 to 750 nm was used. A semi-silvered mirror was used as a reflector. *b*, As *a* but using fluorescence mapping, averaged over 80 trials. Again, the largest 15 signals were marked by triangles. The centres of the two maps of the barrel are superimposed. Note that in this experiment the area of activation with the intrinsic signal is smaller than that seen with the dye signal. (We do not know if the difference is significant because the two measurements were not simultaneous and the fluorescence measurement was done first.) The cortex was stained with 0.1 mg ml⁻¹ RH-795 for 90 min. We used an excitation filter at 530 nm (bandwidth 70 nm), an FT-580 dichroic mirror and an RG610 barrier filter. Common mode noise due to the heartbeat artefact and respiratory movements was removed by subtracting a signal-free trace from the rest of the traces. The duration of the stimulus (*S*), the size of cortical area viewed by each detector, the time scale, and the fractional change in light intensity ($\Delta R/R$ or $\Delta F/F$) reaching the detectors, are marked at the corners of the map. Note the difference in time scale of *a* and *b*. For a discussion of the area of an optically detected barrel relative to its anatomically determined size see the discussion in ref. 5. In similar mapping experiments, using the intrinsic signal, two individual barrels could be resolved if two whiskers were stimulated.

Methods. The optical monitor is described elsewhere^{3,6,16}. The detector amplifiers have a sample-and-hold option that facilitated the high gain d.c. recordings. To minimize the large noise associated with heartbeat, the stimulus and data collection were triggered by the heartbeat, and each stimulus associated trace was subtracted from a trace obtained without stimulus^{5,17}. We used 22 rats, 21 cats and one monkey. Rats were anaesthetized with an intraperitoneal injection of urethane. Cats and monkeys were anaesthetized with continuous infusion of sodium pentothal, paralysed with succinylcholine or curare/gallamine, and artificially respired. The visual stimulus was a square-wave grating, 0.3–1 cycles per degree, drifting at a rate of 2–3 cycles per s. The surface of the cortex was positioned perpendicular to the axis of the microscope objective by a modified stereotaxic apparatus. Most of the light reaching the photodetectors was stray light (80–90%) reflected from the epi-illuminator attachment, the objective lenses, and the coverglass of the chamber attached to the skull. Note that the real fractional change of reflected light from cortex is therefore 5–10-fold larger than that shown on the vertical calibration scales. No attempt was made to minimize the stray reflections because they did not contribute to the noise originating from movements of the cortical surface.

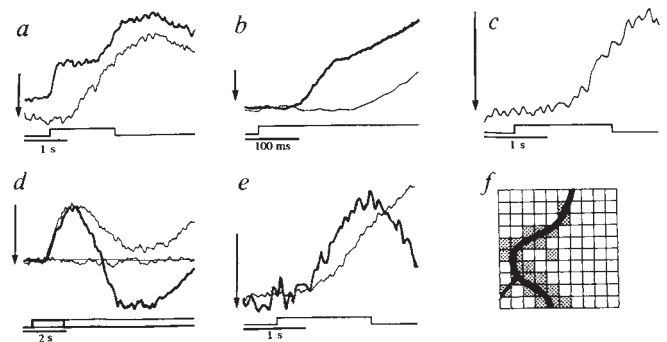


Fig. 2 Characterization of the reflected-light signal in rat, cat and monkey cortex. *a*, Comparison between reflected light signal (fine trace) and the fluorescence signal (bold trace) in the rat somatosensory cortex. Cortical activity was evoked by a 15° deflection of 6 whiskers. The two signals were normalized to the peak. *b*, Comparison between reflected light signal (thin trace) and signal from voltage-sensitive dyes (bold trace), measured from the same cortical area of the cat visual cortex (note the expanded time scale). The amplitudes were normalized so that the late component of the fluorescence would have a similar slope to the light reflection signal. Note the two components of the fluorescence signal in *a* and *b*. (The late component of the fluorescence signal is produced by the intrinsic signal, and not by a change in membrane potential. Cortex stained with non-voltage-sensitive fluorescent dyes, such as fluorescein, gave signals with a time course identical to the intrinsic signal.) *c*, The reflected light signal from the monkey visual cortex shown on a slow time scale. *d*, The relaxation time course of the signal. The shape of the intrinsic signal depended on the duration of the stimulus, having an undershoot to a short (1.5 s) stimulus, but not to a longer (8 s) stimulus. The flat trace was obtained without a stimulus to show the noise associated with the measurements. *e*, Comparison of the time course of the reflected light signal from the barrel (bold line) in the rat cortex and the arterial signal outside the barrel field (fine line). The arterial signals were seen along the full course of an artery (3 mm) which traversed the stimulated barrel. *a–e*, Vertical arrows, size and direction of fractional change in reflected light, $\Delta R/R$ (fine traces; scale in *a–c*, 10^{-4} ; *e*, 2×10^{-4} bold trace; scale in *e*, 10^{-4} ; scale in both traces *d*, 5×10^{-4}) or of fractional change in fluorescence, $\Delta F/F$ (bold traces; scale in *a*, 10^{-3} ; scale in *b*, 2×10^{-4}). *f*, The spatial distribution of the largest reflected light signals monitored at a wavelength of 850 nm (bandwidth of 40 nm), in cat cortex. Detectors that had a signal larger than 50% of the maximal signal are shaded in grey. Their distribution co-localized with an artery (150 μm in diameter) in the field of view even at this wavelength.

duced a large and slow (10 s) undershoot following the peak of the signal. Because of the slow recovery time of the undershoot, to avoid distortion of the signals it is necessary to use inter-stimulus intervals of at least 10 s for visual stimuli lasting 1–2 s.

The cerebral vasculature is the source of one component of the intrinsic signal. In the rat experiment the largest signals were observed along arteries, but not veins, and only when the arteries passed through an activated barrel. The signal associated with arteries was delayed relative to the signals from the active barrel (Fig. 2*e*). The arterial signal was seen at wavelengths ranging from 400 to 1,100 nm (for example, 850 nm, Fig. 2*f*). The other signal components that were useful for functional mapping showed the optimum signal-to-noise ratio at a wavelength range of 610–750 nm. We observed that when using another dye (RH414, ref. 17) the arteries constricted, and only the artery-associated signal disappeared, suggesting that the signal represented changes in blood flow in response to local neuronal activity^{18,19}. This arterial signal could undoubtedly be a useful tool in studying the relationship between neuronal activity and microcirculation but when making functional maps it can introduce distortions. The distortions can be eliminated from the maps with the appropriate analysis, as shown below.

Having optimized the stimulus and recording conditions we obtained optical maps for different stimulus orientations in the

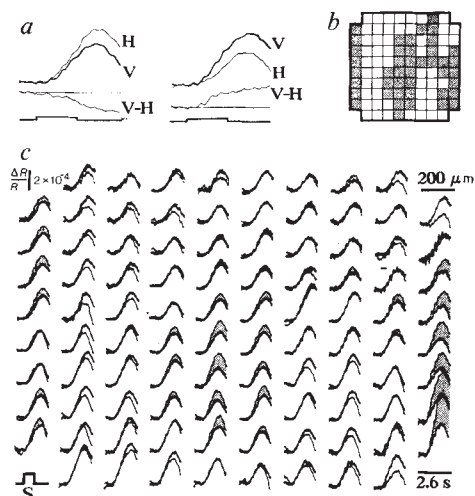


Fig. 3 Visualization of the orientation columns in cat visual cortex. *a*, Demonstration of the signal sensitivity to the orientation of the stimulus. Right traces: the response to vertical drifting gratings (*V*) and to horizontal drifting gratings (*H*). The difference between these two traces is shown in the third trace (*V-H*). An electrode penetration at this site recorded from cells which responded only to vertical gratings. Left traces: as for right, but the optical signals are from detectors that monitored cortical area in which units responded only to horizontal stimuli. *b*, A two-state map of the orientation columns. For each pixel, we integrated the signals for the vertical and horizontal stimuli. Shaded detectors, those giving a larger response to horizontal than to vertical stimuli. *c*, The raw data. Two superimposed patterns of reflected light signals in response to horizontal (fine lines) and vertical (bold lines) drifting gratings. These experiments were interlaced. Positions where the response to horizontal stimuli was larger than the response to vertical stimuli are shaded in grey. The pattern of optical signals may seem puzzling as optical signals are seen everywhere, also in areas where single unit responses were not recorded for a given stimulus orientation²⁰. We attribute this to a component of the signal that is independent of stimulus orientation and lateral spread of activity along dendritic and axonal processes. The signals were averaged over 100 trials. The signal to noise ratio varied from animal to animal and also depended on the physiological state of the animal. In some experiments the pattern became apparent after averaging 10 trials. In other experiments 200 trials were required. Objective criteria such as excessive noise or unusually slow signals could be used to predict whether an optical map is accurate.

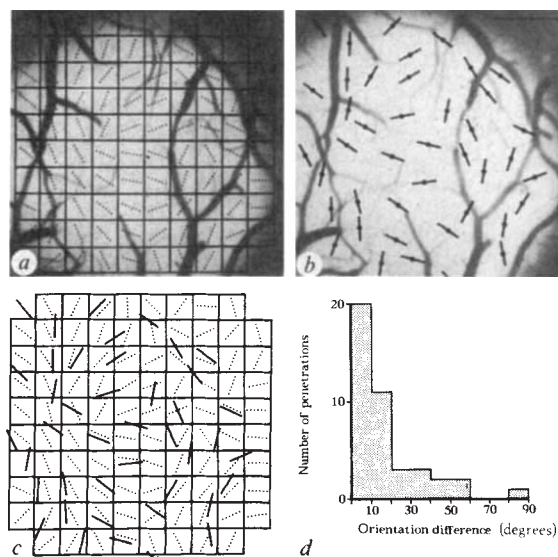


Fig. 4 Electrophysiological confirmation of the optical map of orientation columns. *a*, Optical map of the optimal stimulus orientation, for each cortical site. Since the size of the signal at any given site depends on the inhomogeneous distribution of signal sources as well as the orientation of the stimulus, we took the ratio (rather than the difference) of the integrals of the optical responses to the orthogonal stimulus for each pixel. The ratio approach minimized map distortions produced by inhomogeneous distribution of signal sources (for example, arteries or capillaries). From this ratio we calculated a vector for each pixel. The angle of the optimal stimulus orientation (dotted line) was calculated by vectorial addition of two sets of experiments (see text). The position of the photodiode array (grid) is shown relative to the cortical surface. This map was highly reproducible. A second set of experiments done several hours apart provided an identical orientation map, such that in 95% of the detectors the average difference in the optimal orientation angle determined by the two independent experiments was less than 5°. *b*, The results of the electrophysiological mapping with 42 perpendicular penetrations with tungsten electrodes. The preferred angles of orientation for the single units were determined with a hand-held projector. The location of each penetration was marked on a photograph of the cortical surface. Scale marker, 500 μm . *c*, Superimposed optical (dotted lines) and electrophysiological maps (bold lines). Although there were discrepancies at a few sites, a pixel adjacent to the site of electrode penetration often had an orientation very close to that determined electrophysiologically. Note that large deviations were not associated with the vascular pattern. *d*, Evaluation of the correspondence between the orientations determined electrophysiologically and optically. The histogram shows the deviations between the optimal angle determined by single unit recordings and the optically determined angle at that site or within 100 μm away. Several sources of error are associated with this comparison including the error in the precise marking the sites of the electrode penetrations and the fact that each pixel sums the activity over an area of $200 \times 200 \mu\text{m}$. Note that 74% of the points fell within 20°, the estimated accuracy of the single unit recordings.

cat visual cortex. These maps were compared to maps generated by multiple electrode penetrations. Although signals to both vertical and horizontal stimulus presentations were seen everywhere, in restricted regions of cortex, the optical signals showed a larger response to a horizontal drifting grating (Fig. 3*a*, left, thin trace) than to a vertical grating (thick trace). Other regions showed the opposite behaviour, giving a larger optical signal to the vertical grating (Fig. 3*a*, right). In other cortical sites the optical signals were nearly identical and in those sites single units did not respond to the pair of stimulus orientations used. The difference between the signals to orthogonal stimuli (vertical (*V*)–horizontal (*H*)) showed the signal component that was sensitive to orientation. The spatial distribution of the regions preferring the horizontal stimulus is shown as a two-state coded map in Fig. 3*b*. The raw data are shown in Fig. 3*c*.

We combined the results from vertical and horizontal stimulus presentations (Fig. 3*c*) with results from presentations at orientations of 45° and 135° to obtain a more precise map of the orientation columns. We determined the optimal orientations of each pixel by vector addition (Fig. 4*a*). To confirm the optical map the same area of cortex was mapped, using a double-blind procedure, by making a series of 42 electrode penetrations (Fig. 4*b*). The two maps showed a good correspondence (Fig. 4*c*, *d*). From similar experiments in three different cats we

conclude that the intrinsic signals can be used to map the orientation columns.

Our findings from the intrinsic and dye signals leads to several conclusions concerning their relative merits. The principle advantage of optical mapping using voltage-sensitive dyes is that the time resolution is <1 millisecond, while that of the reflected light signals is in the order of seconds. The better time resolution of the dyes is essential for analysing the cortical circuitry and for understanding the sequential flow of information and its processing at different cortical sites. However, at present the dyes have several shortcomings; some of them may induce pharmacological side effects and cause photodynamic damage upon prolonged or intense illumination, they bleach

and the depth of penetration into the cortex is limited¹⁻⁵. Often, the dye signals are distorted by the intrinsic signals^{6,11}.

A voltage-sensitive dye, NK2367, was used by Blasdel and Salama to study the functional architecture of the monkey striate cortex¹⁴. There are several discrepancies between their work and ours. The first issue is whether their signal came from the dye, as they claim, or from an intrinsic source. The slow time course of the signals seen in both studies suggests that their signal did not come from the dye, as in all other preparations using NK2367 the signal included a fast component²¹⁻²⁴ even if a slow glial signal was also present²⁵. Furthermore, the standard evidence to attribute a signal to a dye is the observation of different signals at different wavelengths, and this evidence was not provided. Another issue is the usefulness of the intrinsic signal for mapping. Although Blasdel and Salama observed an intrinsic signal, they did not find it useful for functional mapping. Finally, there is the question of distortion of maps by the cortical vasculature. As we show, the blood vessel signals only appear when the cortex is active (that is, in the presence of a stimulus). The fact that no pattern is seen in the control experiment, in the absence of a visual stimulus leads Blasdel and Salama to conclude that the vascular pattern does not contribute to the maps. In their work, however, the vascular pattern is visible in the functional maps. Although we see arterial signals in our raw data, there is no arterial contribution to the functional maps (for example, Fig. 4), possibly because of the method of analysis and data acquisition, the low spatial resolution of the recording system or a lower noise level.

The exact origin of the multiple components of the reflected light signal will determine the ultimate spatial resolution and remains to be investigated. In addition to the arterial component, which probably reflects changes in blood volume or flow, a second component probably originated from local activity light scattering signals as described in previous reports^{6-11,26}. The third component may originate from local changes in the absorp-

tion associated with the conversion of oxyhaemoglobin to haemoglobin²⁷⁻²⁹ in the capillaries, in response to metabolic demands. In fact we have recently found that this component is a significant constituent of the intrinsic signal.

We thank Drs L. Katz and E. Kaplan for their useful comments. Supported by NIH grants NS14716 and EY05253, NSF grants BNS9318794 and 8351738 and grant AFOSR-85-0301 from the US Air Force and by IBM.

Received 21 May; accepted 24 September 1986.

- Salzberg, B. M., Davila, H. V. & Cohen, L. B. *Nature* **246**, 508-509 (1973).
- Salzberg, B. M. in *Current Methods in Cellular Neurobiology* (ed. Barker, J. L.) 139-187 (Wiley, New York, 1983).
- Grinvald, A. *Rev. Neurosci.* **8**, 263-305 (1985).
- Cohen, L. B. & Leshner, S. L. in *Optical Methods in Cell Physiology* (eds DeWeer, P. & Salzberg, B. M.) 71-99 (Wiley, New York, 1986).
- Orbach, H. S., Cohen, L. B. & Grinvald, A. *J. Neurosci.* **5**, 1553-1564 (1985).
- Grinvald, A., Manker, A. & Segal, M. *J. Physiol., Lond.* **333**, 269-291 (1982).
- Hill, D. K. & Keynes, R. D. *J. Physiol., Lond.* **108**, 278-281 (1949).
- Cohen, L. B., Keynes, R. D. & Hille, B. *Nature* **218**, 438-441 (1968).
- Tasaki, I., Watanabe, A., Sandlin, R. & Carnay, L. *Proc. natn. Acad. Sci. U.S.A.* **61**, 883-888 (1968).
- Lipton, P. *J. Physiol., Lond.* **231**, 365-383 (1973).
- Salzberg, B. M., Obaid, A. L. & Gainer, H. *J. gen. Physiol.* **86**, 395-411 (1985).
- Chance, B., Cohen, P., Jobsis, F. & Schoener, B. *Science* **137**, 499-508 (1962).
- Jobsis, F. F., Keizer, J. H., LaManna, J. C. & Rosenthal, M. *J. appl. Physiol.* **43**, 858-872 (1977).
- Blasdel, G. G. & Salama, G. *Nature* **321**, 579-585 (1986).
- Woolsey, T. A. & Van der Loos, H. *Brain Res.* **17**, 205-242 (1970).
- Grinvald, A., Cohen, L. B., Leshner, S. & Boyle, M. B. *J. Neurophysiol.* **45**, 829-840 (1981).
- Grinvald, A., Anglister, L., Freeman, J. A., Hildesheim, R. & Manker, A. *Nature* **308**, 848-850 (1984).
- Roy, C. S. & Sherrington, C. S. *J. Physiol., Lond.* **11**, 85-108 (1890).
- Lassen, N. A. & Ingvar, D. H. *Experientia* **17**, 42-43 (1961).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106-154 (1962).
- Salzberg, B. M., Grinvald, A., Cohen, L. B., Davila, H. V. & Ross, W. N. *J. Neurophysiol.* **40**, 1281-1291 (1977).
- Grinvald, A., Salzberg, B. M. & Cohen, L. B. *Nature* **268**, 140-142 (1977).
- Sensenbrenner, D. M. & Salzberg, B. M. *Science* **208**, 1267-1271 (1980).
- Ross, W. N. & Krauthamer, V. *J. Neurosci.* **4**, 659-672 (1984).
- Konnerth, A. & Orkand, R. K. *Neurosci. Lett.* **66**, 49-54 (1986).
- Cohen, L. B. *Physiol. Rev.* **353**, 373-418 (1973).
- Millikan, G. A. *Proc. R. Soc. B123*, 218-241 (1937).
- Jobsis, F. F. *Science* **198**, 1264-1267 (1977).
- Moos, G. A., Osypka, P., Strum, R. E. & Wood, E. H. *Cardiovasc. Res.* **2**, 199-209 (1968).

Control of malarial invasion by phosphorylation of the host cell membrane cytoskeleton

K. Rangachari, A. Dluzewski, R. J. M. Wilson* & W. B. Gratzer

Medical Research Council Cell Biophysics Unit, King's College, Drury Lane, London WC2B 5RL, UK

* National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK

It has been shown that the entry of the malaria parasite into the red blood cell requires the presence of ATP in the host cell cytoplasm^{1,2}. In red blood cell ghosts that contain no ATP the receptor on the extracellular surface remains in place and parasites will bind to the membrane, but will not enter³. ATP is thus necessary for one of the steps⁴ in the invasion sequence that follows recognition and attachment. The process of entry appears to involve the active participation of the host cell membrane cytoskeleton⁵. We have suggested² that the function of the intracellular ATP may be to regulate phosphorylation of the cytoskeleton. We now present evidence that the activity of the membrane-associated cyclic AMP-independent kinase of the red blood cell is inseparable from invasion; the active substrate may be spectrin.

Kinases are found in both the membrane and cytoplasm of the human red cell, and there is extensive turnover of phosphoryl groups on cytoskeletal proteins. Spectrin is phosphorylated in the cell by a cAMP-independent (casein) kinase at four sites, clustered at one end of the smaller (β) subunit⁶. Early evidence that phosphorylation at these sites was linked to metabolic shape

control⁷ did not survive closer scrutiny^{8,9} and there is no known function for the phosphorylation. However, it does appear that the phosphorylation state of spectrin can be perturbed by external stimuli^{10,11} and migration of intramembrane particles may also be linked to spectrin phosphorylation^{10,12}. Rearrangement of intramembrane particles appears to occur during entry of the malarial parasite. A bare area of membrane develops in the region of contact¹³ and a dense zone, consisting almost certainly of clustered intramembrane particles, can be discerned in the electron microscope at the parasite-host cell junction when the merozoite begins to move into the cell⁴. To establish whether invasion depends on phosphorylation of spectrin (or of other membrane proteins) we used our earlier observation that malaria parasites can, under defined conditions, be made to enter resealed red blood cell ghosts¹⁴ and the discovery that the cAMP-independent kinase can be extracted from the membrane by a medium of high ionic strength⁹.

We found that fresh ghosts hypotonically lysed in the presence of ATP could be subjected to one buffer wash before resealing in the presence of the original haemolysate and 1 mM ATP, without loss of susceptibility to invasion. Freshly lysed ghosts were exposed to a buffer containing 0.65 M potassium chloride to extract the kinase⁹, then recovered by centrifugation and resealed with haemolysate and ATP. Inoculation of these ghosts with purified schizonts of *Plasmodium falciparum* revealed drastically reduced susceptibility to invasion (Table 1).

A crude preparation of the cAMP-independent kinase can be made from a high-salt extract of membranes¹⁵. When this was added (~100 μ g per ml final protein) to the haemolysate, invasion of salt-extracted ghosts was substantially restored. This enzyme preparation contains a wide range of membrane proteins, of which the kinase makes up a very small part. It is thus open to question whether it was kinase, rather than some other

Table 1 Effect of cAMP-independent kinase on invasion of red cells by *P. falciparum*

Sample	Invasion ratio*								Average
	1	2	3	4	5	6	7	8	
Untreated	1.3	3.0	1.5	2.0	2.7	0.9	0.7	—	1.73 ± 0.36
Kinase extracted	0.6	0.4	0.3	0.9	0.8	0.1	0	0.2	0.41 ± 0.12
Kinase replaced	1.9	1.2	1.0	1.8	2.0	0.9	0.8	1.0	1.33 ± 0.19

* Ratio of fraction of parasitized cells after one generation (~24 h) to that in the culture after addition of inoculum. Numbers 1-8 are separate experiments with different cultures: the kinase in experiments 1-5 was the purified enzyme, and in 6-8 was the crude preparation. Averages and standard mean errors are given.

Ghosts were prepared by hypotonic lysis of washed human red cells in 5 mM phosphate, 2 mM magnesium chloride, 2 mM ATP, pH 7.4. The membranes were collected by centrifugation and extracted with four volumes of 0.65 M potassium chloride, 0.1 mM EDTA, 0.25 mM dithiothreitol, pH 7.4, on ice for 15 min to remove kinase⁹. The pellet was washed once with 0.14 M potassium chloride, 2 mM ATP. Extracted and untreated ghosts were resealed by incubation at 37 °C with a diluted haemolysate (usually 1:8 with isotonic saline, containing 2 mM magnesium chloride, 2 mM ATP) and a kinase preparation as required. The total added protein, estimated colorimetrically¹⁹, was ~100 and <0.4 µg ml⁻¹, in the case of the crude and purified kinase, respectively. The resealed ghosts were twice washed with culture medium (RPMI 1640+10% human serum) and suspended at 50% haematocrit. Schizonts of *P. falciparum* were purified from *in vitro* cultures^{17,18} and were added to the ghosts. The extent of parasite invasion was determined after ~24 h by counting the number of new parasites in 1,000 cells in thin smears, stained with Giemsa's solution. A crude kinase preparation was obtained by extraction of cytoskeleton-depleted red cell membranes with a medium containing 0.5 M NaCl as described¹⁵. The extract was fractionated, after precipitation with ammonium sulphate, by two successive elutions from Sephadex G-100, the first in 0.1 M, the second in 0.5 M NaCl. Two zones of activity eluted from the second column, the first with spectrin in the void volume, the second (free kinase) later. Activity was assayed with [γ -³²P]ATP and casein¹⁵. The peak active fractions were concentrated by vacuum dialysis.

membrane constituent, that was responsible for restoration of capacity for invasion. A further fractionation of the kinase preparation was therefore undertaken using the chromatographic procedure described in ref. 16. The activity elutes at high ionic strength from Sephadex G-100, in two regions, one of high relative molecular mass (M_r), which also contains extracted spectrin, the other of M_r ~34,000. The latter is free kinase, and contains little contaminating protein. The total protein added back in this form to the ghosts before resealing had a concentration of <0.4 µg ml⁻¹. Phosphorylation of membrane proteins was measured by autoradiography of electrophoretic gels of the total solubilized membrane protein after incubation of hypotonic ghosts with ³²P-labelled ATP. The single high-salt extraction (Table 2) removed about half of the endogenous cAMP-independent kinase activity. A second extraction led to extensive vesiculation and misshapen residual ghosts, and was therefore not used. Nevertheless, ghosts in which spectrin phosphorylation was reduced by up to 50% showed consistent reduction of invasion by *P. falciparum* (60-80%; Table 1). Addition of the kinase restored invasion to essentially the original level. Kinase which did not cause phosphorylation (inactivated by some days' storage) also did not restore invasion. The requirement for ATP is absolute; non-hydrolysable analogues are ineffective².

Table 2 Effect of salt-extraction and replacement of kinase on phosphorylation of spectrin in ghosts

Sample	Relative areas under zone profiles		
	Protein stain	Autoradiograph	Relative ³² P incorporation
Untreated	1.00	1.00	1.00
Salt-extracted	1.15	0.60	0.52
Kinase replaced	1.03	1.27	1.23

Unsealed ghosts were incubated with 1 mM [γ -³²P]ATP (20-40 c.p.m. pmole⁻¹) in hypotonic buffer (as in Table 1) on ice for 20 min. The ghosts were made isotonic and resealed at 37 °C, washed twice in isotonic saline and lysed in the hypotonic buffer. Membranes were collected by centrifugation, dissolved in 1% SDS in boiling water and subjected to gel electrophoresis and autoradiography². The autoradiograph and Coomassie blue-stained gel were both evaluated by densitometry. The relative degree of phosphorylation of spectrin is expressed as the ratio of the areas under the zone profiles, taking that for unextracted membranes to be unity. Some 80-90% of the label was found in the spectrin.

The presence of cAMP in resealed ghosts had no influence on invasion efficiency, which is therefore not related to cAMP-dependent kinase(s). We conclude that a substrate for the cAMP-independent membrane kinase must be phosphorylated if malarial invasion is to occur. This is presumably the basis of the requirement for intracellular ATP^{1,2}.

Spectrin is by far the most abundantly labelled protein under the conditions of these experiments. On addition of kinase to the extracted cells the enhanced incorporation of label is about 80-90% in spectrin. Given the earlier evidence for direct involvement of the host cell spectrin in the invasion process^{1,14}, it is a reasonable conjecture (although it cannot be regarded as proven) that spectrin is the invasion-linked substrate for the cAMP-independent kinase.

This investigation received financial support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. We thank Dr J. C. Pinder for discussion and advice on kinase preparation.

Received 27 February; accepted 30 September 1986.

- Olson, J. A. & Kilejian, A. *J. Cell Biol.* **95**, 757-762 (1982).
- Dluzewski, A. R., Rangachari, K., Wilson, R. J. M. & Gratzer, W. B. *Molec. biochem. Parasitol.* **9**, 145-160 (1983).
- Miller, L. H. *Bull. WHO* **55**, 157-162 (1977).
- Aikawa, M., Miller, L. H., Johnson, J. & Rabbege, J. *J. Cell Biol.* **77**, 72-82 (1978).
- Dluzewski, A. R., Rangachari, K., Wilson, R. J. M. & Gratzer, W. B. *Br. J. Haematol.* **55**, 629-637 (1983).
- Harris, H. W. & Lux, S. E. *J. biol. Chem.* **255**, 11512-11520 (1980).
- Birchmeier, W. & Singer, S. J. *J. Cell Biol.* **73**, 647-659 (1977).
- Anderson, J. & Tyler, J. M. *J. biol. Chem.* **255**, 1259-1265 (1980).
- Patel, V. P. & Fairbanks, G. *J. Cell Biol.* **88**, 430-449 (1981).
- Loyter, A. A., Ben-Zaquen, R., Marash, R. & Milner, Y. *Biochemistry* **16**, 3903-3909 (1977).
- Nelson, M. J. & Huestis, W. H. *Biochim. biophys. Acta* **600**, 398-405 (1980).
- Lalazar, A. & Loyter, A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 318-321 (1979).
- McLaren, D. J., Bannister, L. H., Trigg, P. I. & Butcher, G. A. *Parasitology* **79**, 125-139 (1979).
- Dluzewski, A. R., Rangachari, K., Wilson, R. J. M. & Gratzer, W. B. *Br. J. Haematol.* **49**, 97-101 (1981).
- Hosey, M. M. & Tao, M. *Biochim. biophys. Acta* **482**, 348-357 (1977).
- Tao, M., Conway, R. & Cheta, S. *J. biol. Chem.* **255**, 2563-2568 (1980).
- Trager, W. & Jensen, J. B. *Science* **193**, 673-675 (1976).
- Dluzewski, A. R., Ling, I. T., Rangachari, K., Bates, P. & Wilson, R. J. M. *Trans. R. Soc. trop. Med. Hyg.* **78**, 622-637 (1984).
- Bradford, M. M. *Analyt. Biochem.* **72**, 248-254 (1976).
- Pinder, J. C., Dhermy, D., Baines, A. J., Lux, S. E. & Gratzer, W. B. *Br. J. Haematol.* **56**, 455-463 (1983).

Plant chitinases are potent inhibitors of fungal growth

Angela Schlumbaum, Felix Mauch,
Urs Vögeli & Thomas Boller

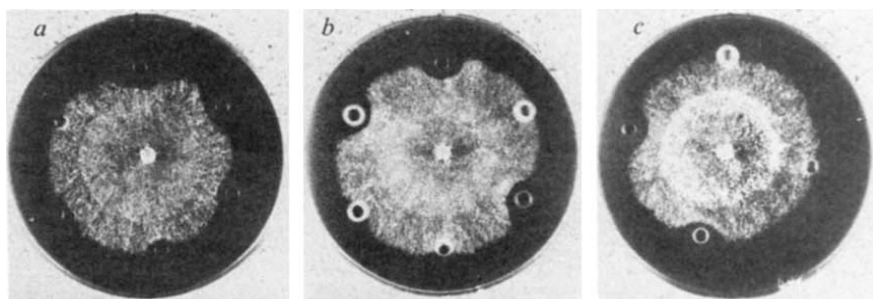
Abteilung Pflanzenphysiologie, Botanisches Institut,
Universität Basel, Hebelstrasse 1, CH-4056 Basel, Switzerland

The antimicrobial arsenal of plants is thought to consist mainly of secondary metabolites, among which the phytoalexins are the best-studied¹⁻³. But plants may also possess antimicrobial proteins^{4,5}: it has been reported that wheat-germ agglutinin, a chitin-binding lectin from wheat embryos, inhibits growth of the fungus *Trichoderma viride*⁴. This has led to the notion that plant lectins, with their intriguing biochemical similarity to animal antibodies, have an antibody-like antimicrobial function^{4,6,7}. We report here that the main proteinaceous inhibitor of fungal growth in bean leaves is chitinase, an enzyme that can be induced by the plant hormone ethylene, or by pathogen attack. Among commercial preparations of purified chitin-binding lectins (from wheat germ, tomato, potato, pokeweed and gorse), only those containing contaminating chitinase activity inhibit fungal growth. Our data indicate that plant chitinases, but not chitin-binding lectins, are important antifungal proteins in plants.

The resistance of plants to potentially pathogenic microorganisms is based on multiple biochemical factors¹. Particularly important are the phytoalexins, antimicrobial micromolecules that are not present in healthy plants but accumulate in response

Fig. 1 Growth inhibition of *T. viride* in response to various protein preparations. **a**, Purified bean chitinase¹³. Wells contain 20 μ l of the following solutions in μ g ml⁻¹ (from top, in clockwise direction): 600; 150; 37; 9; 2; 0.6. **b**, Purified chitinase and crude protein extracts from primary bean leaves. Wells contain 20 μ l of the following solutions in μ g ml⁻¹ (from top, in clockwise direction): Purified bean chitinase (130); purified bean chitinase (130) in the presence of antiserum against chitinase; crude protein (12 mg ml⁻¹) from leaves treated with 10 nl ml⁻¹ ethylene for 48 h; crude protein (12 mg ml⁻¹) from control leaves; crude protein (12 mg ml⁻¹) from ethylene-treated leaves in the presence of antiserum against chitinase; and crude protein (12 mg ml⁻¹) from ethylene-treated leaves in the presence of preimmune serum. The chitinase activities, determined after the addition of the sera, corresponded to the following concentrations of bean chitinase in μ g ml⁻¹ (from top, in clockwise direction): 130; 0.1; 130; 1.4; 1.3; 130. **c**, Purified bean chitinase and wheat germ agglutinin (WGA). Wells contain 20 μ l of the following solutions (from top, in clockwise direction): commercial wheat germ agglutinin (25 mg ml⁻¹); commercial wheat germ agglutinin (5 mg ml⁻¹); wheat germ agglutinin purified by affinity chromatography on regenerated chitin, 3.5 mg ml⁻¹ (with a chitinase activity corresponding to 115 μ g ml⁻¹ bean chitinase); and bean chitinase, 100 μ g ml⁻¹.

Methods. *T. viride* (a gift of Dr Dahmen) was grown on Petri dishes containing 2% malt extract agar at the bottom and 1.5% malt extract agar at the top. One day after inoculation, when the colony diameter was 3–4 cm, wells punched into the 1.5% agar layer 1 cm from the growing front were filled with 20 μ l of the various preparations. The plates were photographed 8–12 h later. The crude protein extracts were prepared using 0.1 M potassium phosphate buffer, pH 7.0 (4 ml per g fresh weight). After centrifugation (10 min at 10,000 g), the supernatant was precipitated overnight with ammonium sulphate at 97% saturation, taken up in half the extraction volume and dialysed twice against 0.01 M potassium phosphate buffer, pH 7.0. The extracts were filtered through a 0.22 μ m membrane filter (Sartorius, type SM 16534), lyophilized, and taken up in water. Chitinase and protein were measured as previously described¹³.



to microbial infection^{1–3}. Many of these phytoalexins inhibit growth of fungi at concentrations of 10–50 μ g ml⁻¹. In addition, plants may possess proteins with antibiotic activity^{4,5}. This was first demonstrated when Mirelman *et al.*⁴ found that wheat-germ agglutinin (WGA), a chitin-binding lectin, inhibited growth of the fungus *T. viride* at concentrations of ≥ 600 –1200 μ g ml⁻¹. Several lectins have since been reported to inhibit fungal growth under similar conditions^{8,9}. Plant lectins that bind to fungal cell walls are therefore thought to have an antifungal role^{5,6,9}.

Another group of plant proteins with potential antifungal functions are the hydrolases that can attack the main components of fungal cell walls, chitin and β -1,3-glucan^{1,10}. We have studied chitinase, an enzyme with no known endogenous substrate in higher plants^{10,13}, which has been purified from various plant tissues, including wheat germ¹¹, tomato leaves¹² and bean leaves¹³. This enzyme is strongly induced in response to the plant hormone, ethylene¹³, and in response to infection by pathogens^{12,14,15}. Purified plant chitinase readily liberates chitin oligomers from purified fungal cell walls^{12,13} and acts as a lysozyme against bacterial cell walls^{10,13}. This is indirect evidence for antimicrobial activity of the enzyme.

To examine antifungal activity directly, we tested the effect of purified bean chitinase on growth of *T. viride*, using the method of Mirelman *et al.*⁴. We observed a marked concentration-dependent inhibition of fungal growth around the wells containing chitinase (Fig. 1a). The smallest concentration of chitinase that resulted in a distinct inhibition zone was 2 μ g ml⁻¹. This means that chitinase is 300 times as potent as the WGA preparation of Mirelman *et al.*⁴.

Boiled chitinase had no effect on fungal growth. Most plant chitinases are basic proteins¹⁰, but another basic protein, cytochrome c, did not inhibit fungal growth in the bioassay at concentrations up to 10 mg ml⁻¹ (data not shown). When antibodies were raised against purified bean chitinase in a rabbit, the antiserum was found to block chitinase activity. It also blocked the inhibitory effect of chitinase on fungal growth (Fig. 1b) whereas preimmune serum of the same rabbit had no such effect.

We tested crude protein extracts from bean leaves in the fungal growth assay, using a protein concentration (12 mg ml⁻¹) equivalent to that in leaves (Fig. 1b). Protein extracts from ethylene-treated leaves produced much larger inhibition zones than extracts from control leaves. This correlates well with strong

induction of chitinase by ethylene¹³. The crude extracts from ethylene-treated leaves had a chitinase activity equivalent to 130 μ g ml⁻¹ pure chitinase. The same amount of pure chitinase produced an inhibition zone of a similar size. The inhibitory effect of the crude extract was almost completely overcome by the addition of antiserum against chitinase, whereas preimmune serum did not affect it (Fig. 1b). These data indicate that chitinase is the main inhibitor of fungal growth in crude protein preparations from bean leaves.

We have also examined bean leaves from plants infected with bean rust, *Uromyces phaseoli*, using the bean cultivar Warox, which is hypersensitively resistant to the pathogen used. We found that chitinase was about tenfold induced in response to infection. Crude extracts from infected leaves produced larger inhibition zones than did extracts from uninfected control leaves; again, adding the antiserum against bean chitinase prevented the formation of an inhibition zone (Table 1).

Table 1 Inhibition of fungal growth by crude extracts from rust-infected and uninfected bean leaves

Test Preparation	Protein mg ml ⁻¹	Chitinase μ g ml ⁻¹	Inhibition of Fungal Growth
Extract from uninfected leaves	7.0 3.5	3.3 1.6	+ –
Extract from infected leaves	6.6 3.3 1.7	30 15 7.5	+++ +++ ++
Extract from infected leaves and antiserum	6.6	<0.05	–

Bean seedlings (c. Warox, supplied by van Waveren Pflanzenzucht, Rosdorf, FRG) were sprayed 10–14 days after germination with 1% Tween-80 (controls) or with 1% Tween-80 containing 6×10^4 spores of bean rust per ml (*U. phaseoli*, supplied by Dr U. Gisi). The plants were kept for 48 h at 100% humidity and further incubated in a greenhouse for 12 days. Protein extracts from primary leaves were prepared as described in the legend to Fig. 1. Chitinase activity¹³ is expressed in equivalents of pure bean chitinase¹⁵. The effect on growth of *T. viride* (tested with 20 μ l of each preparation as in Fig. 1) is expressed in qualitative terms: +++ stands for strong inhibition, + for just detectable inhibition.

Table 2 Effect of chitin-binding plant lectins on fungal growth

Lectin preparation	Protein mg ml ⁻¹	Titre µg ml ⁻¹	Chitinase µg ml ⁻¹	Inhibition of fungal growth
Tomato lectin*	2	<0.5	13	+++
Potato lectin*	2	0.9	1.7	+
Pokeweed lectin†	5	30	0.7	++
Pokeweed lectin*	3	0.2	<0.05	-
Gorse lectin				
UEA II*	2	60	0.06	-

The titre for the agglutination of trypsinized red blood cells and the chitinase activity¹³, expressed in equivalents of pure bean chitinase¹⁵, were determined for each lectin preparation. The effect on growth of *T. viride* (tested with 20 µl of each preparation as in Fig. 1) is expressed in qualitative terms as in Table 1.

* Obtained from Sigma.

† Obtained from Serva.

To reexamine the antifungal effect of WGA⁴, we verified that purified WGA (Serva) agglutinated trypsinized red blood cells (titre: 2.4 µg ml⁻¹) and found that it contained no measurable chitinase activity. This WGA preparation did not inhibit growth of *T. viride* even at concentrations of 25 mg ml⁻¹ (Fig. 1c). Preparations of WGA from Sigma, Calbiochem and Pharmacia were similarly free of contaminating chitinase and did not inhibit fungal growth.

Mirelman *et al.*⁴ purified WGA by affinity chromatography on sepharose-2-acetamido-N-(5-aminocaproyl)-2-deoxy-β-D-glucopyranosylamine. However, affinity chromatography of WGA may result in copurification of WGA and chitinase¹¹. WGA prepared by affinity chromatography on regenerated chitin^{11,16} (a readily available affinity matrix) agglutinated trypsinized red blood cells with a titre of 3.3 µg ml⁻¹ and contained contaminating chitinase activity equivalent to ~40 µg bean chitinase per mg protein. (Wheat-germ chitinase had previously been purified from a similar affinity-purified WGA preparation¹¹.) This chitinase-contaminated WGA preparation strongly inhibited growth of *T. viride* at a concentration of 3.5 mg ml⁻¹ (Fig. 1c). The antiserum against bean chitinase did not inhibit the chitinase activity in WGA and did not reduce the size of the inhibition zone caused by the contaminated WGA. When pure bean chitinase was added to chitinase-free commercial WGA, the combination inhibited fungal growth in the same way as affinity-purified WGA (data not shown). Thus, the inhibitory effect of WGA observed by Mirelman *et al.*⁴ may have been caused by contamination with chitinase.

We have also tested the effect of several other commercial lectins with a specificity for N,N'-diacetylchitobiose or N,N',N''-triacylchitotriose (Table 2). The tomato and the potato lectin preparations contained contaminating chitinase activity. Both inhibited fungal growth. The chitinases present in these preparations were found to be serologically related to bean chitinase as the antiserum against bean chitinase totally inhibited the chitinase activity of both. The antiserum also suppressed the growth-inhibiting activity of the lectin preparations completely (data not shown). We obtained two different preparations of pokeweed (*Phytolacca americana*) lectin. One preparation (Serva), had lower lectin activity than the other. It was contaminated with chitinase and inhibited fungal growth (Table 2). The other preparation (Sigma), although more active as a lectin, did not contain chitinase and did not inhibit fungal growth. A gorse agglutinin, UEA II, contained almost no chitinase activity and did not inhibit fungal growth. Collectively, these data show that preparations of chitin-binding lectins inhibit growth of *T. viride* only when they are contaminated by chitinase. Of further commercial lectin preparations with other specificities examined (concanavalin A, bean phytohemagglutinin A, peanut agglutinin, osage orange lectin), none was contaminated by chitinase, and none inhibited fungal growth.

We have demonstrated directly that plant chitinases are highly potent growth inhibitors of *T. viride*. It is likely that the effect is attributable to enzymatic attack on newly formed chitin in the growing hyphal tips¹¹ disturbing the balance between synthesis and hydrolysis required for tip growth¹⁷. It is intriguing that chitinase and β-1, 3-glucanase are coordinately induced by ethylene or by pathogen attack in plants^{10,14}. Chitinase alone, at concentrations of 2 µg ml⁻¹, is sufficient to stop growth of *Trichoderma viride*, a non-pathogenic test fungus. Inhibition of several other fungi, however, requires the presence of a combination of chitinase and β-1, 3-glucanase at similar concentrations (F.M., B. Mauch and T.B., in preparation). On a concentration basis, antimicrobial plant hydrolases, either alone or in combination, are inhibitors of fungal growth at least as potent as the phytoalexins. They may be similarly important in the defences of plants against pathogens.

This work was supported by grant 3.400.0.83 of the Swiss National Science Foundation.

Received 22 July 1986; accepted 30 September 1986.

1. Bell, A. A. *Rev. Pl. Physiol.* **32**, 22-81 (1981).
2. Bailey, J. A. & Mansfield J. W. (eds) *Phytoalexins* (Halsted, Wiley, New York, 1982).
3. Darvill, A. G. & Albersheim, P. A. *Rev. Pl. Physiol.* **35**, 243-275 (1984).
4. Mirelman, D., Galun, E., Sharon, N. & Lotan, R. *Nature* **256**, 414-416 (1975).
5. Roberts, W. K. & Selitrennikoff, C. P. *Biochim. biophys. Acta* **880**, 161-170 (1986).
6. Sequeira, L. A. *Rev. Phytopathol.* **16**, 453-481 (1978).
7. Etzler, M. E. A. *Rev. Pl. Physiol.* **36**, 209-234 (1985).
8. Barkai-Golan, R., Mirelman, D. & Sharon, N. *Archs. Microbiol.* **116**, 119-124 (1978).
9. Brambl, R. & Gade, W. *Physiologia Pl.* **64**, 402-408 (1985).
10. Boller, T. in *Cellular and Molecular Biology of Plant Stress* (eds J. L. Key & T. Kosuge), 247-262 (Liss, New York, 1985).
11. Molano, J., Polacheck, I., Duran, A. & Cabib, E. *J. biol. Chem.* **254**, 4901-4907 (1979).
12. Pegg, G. F. & Young, D. H. *Physiol. Pl. Pathol.* **21**, 389-409 (1982).
13. Boller, T., Gehri, A., Mauch, F. & Vögeli, U. *Planta* **157**, 22-31 (1983).
14. Mauch, F., Hadwiger, L. A. & Boller, T. *Pl. Physiol.* **76**, 607-611 (1984).
15. Métraux, J. P. & Boller, T. *Physiol. molec. Pl. Pathol.* **28**, 161-169 (1986).
16. Bloch, R. & Burger, M. M. *Biochem. biophys. Res. Commun.* **58**, 13-19 (1974).
17. Farkas, V. *Microbiol. Rev.* **43**, 117-144 (1979).

Sulphonolipids are molecular determinants of gliding motility

Darren R. Abbanat, Edward R. Leadbetter, Walter Godchaux III & Alan Escher

Department of Molecular and Cell Biology, U-131, University of Connecticut, Storrs, Connecticut 06268, USA

A variety of bacteria¹ that possess no obvious locomotor organelles are, nonetheless, able to translocate on solid surfaces (but not through liquids). In no case has the mechanism of this 'gliding' motility been elucidated². Our recent discovery that unusual sulphonolipids^{3,4} (Fig. 1) are major cell-envelope components peculiar to the *Cytophaga-Flexibacter* group of gliding bacteria⁵ led us to examine whether these lipids are important in this motility. Mutants deficient in both gliding and sulphonolipid synthesis were isolated from mutagenized cultures of *Cytophaga johnsonae* (our laboratory strain, originally ATCC 17061). In some of these mutants, restoration of sulphonolipid content through provision of a specific biosynthetic precursor resulted in recovery of the ability to glide. The sulphonolipids are the first molecules shown to be specifically required for gliding motility (surface translocation).

Cultures ($A_{650}=0.8$) mutagenized with *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (0.3 mg ml⁻¹) for 60 min were diluted and plated on LTY agar (Fig. 2 legend); 48 spreading-deficient clones were selected on the basis of colony morphology. The colonies formed by gliding cells expand rapidly, are thin and have finger-like projections at their edges, whereas those of nonmotile cells are compact, rounded and have sharply demarcated edges (Fig. 2). In minimal liquid medium (fructose plus inorganic salts, with sulphate as the sole sulphur source), all the mutant strains considered below grew with the same doubling time ($\pm 5\%$) as

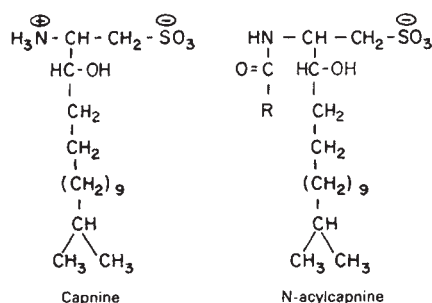


Fig. 1 Structures of the sulphonolipids. R-CO- represents one of a variety of long-chain fatty acids⁴. N-acylcapnine is the predominant species found in the parent strain of *C. johnsonae*; capnine is apparently a precursor of the acylated form.

the parent strain, so their abnormal colony morphology was not the result of growth defects. These spreading-deficient strains were examined for sulphonolipid content. Cells were grown in minimal medium containing $\text{Na}_2^{35}\text{SO}_4$ and lipid radioactivity was measured after separation by thin-layer chromatography (methodological details are given in ref. 5). Four strains were found to be deficient in some aspect of their sulphonolipid content.

Mutant 21A2I contained less than 2% of the amount of sulphonolipid found in the parent strain (Table 1) and was a virtual nonglider, as judged not only by the failure of its colonies to spread (Fig. 2, top row) but also by direct observation using the phase-contrast microscope. Gliding cells form 'rafts'—groups of cells gliding together—that can be observed when they detach from the colony edge. In the parent strain these rafts moved at an average rate of about $25 \mu\text{m min}^{-1}$ (Table 1). In contrast, the edges of colonies of mutant 21A2I were smooth and no gliding movement of cells could be observed.

We have shown⁶ that L-cysteate is the immediate precursor of carbons 1 and 2, the sulphonate group and (probably) the amino group of capnine (Fig. 1). Cysteate is thought to condense with an activated fatty acid to form, ultimately, capnine⁶, in a manner analogous to the biosynthesis of sphingosine from serine and a fatty acyl coenzyme A⁷; capnine is then N-acylated. Provision of L-cysteate in the growth medium restored the sulphonolipid content of mutant 21A2I to normal (Table 1); as is the case in the parent strain⁵, virtually all the sulphonolipid thus formed was in the form of N-acylcapnine. Provision of L-cysteate also restored gliding motility; colony spreading became normal (Fig. 2, bottom row) and rafts were formed and moved at about the same speed as those of the parent strain (Table 1). Evidently mutant 21A2I is unable to form cysteate, an obligatory precursor of capnine, from the standard sulphur sources provided in these experiments but when given cysteate it can synthesize sulphonolipid and therefore can glide. A second mutant, number 42, is apparently a 'leaky' version of 21A2I. It contains detectable sulphonolipid but only 13% of normal levels; its colony-spreading and raft-movement rates are also low (70 and 30% of the parent-strain rates, respectively) and L-cysteate restores gliding ability to normal.

The effects of cysteate are stereospecific; D-cysteate, which does not compete with L-cysteate for incorporation into capnine (unpublished observations), did not restore gliding motility (Fig. 2, middle row). Hence, the effect of L-cysteate on gliding must be a biological one and not some gross physical effect of a sulphonate in the medium. A number of other compounds tested, including L-homocysteate, taurine, L- and D-cystine, L-cysteine sulphinic acid, L-methionine, L-djenkolate and 2-mercaptoethanesulphonic acid (coenzyme M), proved equally ineffective in restoring gliding. We also note that *C. johnsonae* cannot use cysteate as the sole sulphur source (unpublished observation) and that sulphur and carbon from exogenous L-cysteate are incorporated into sulphonolipids but not into pro-

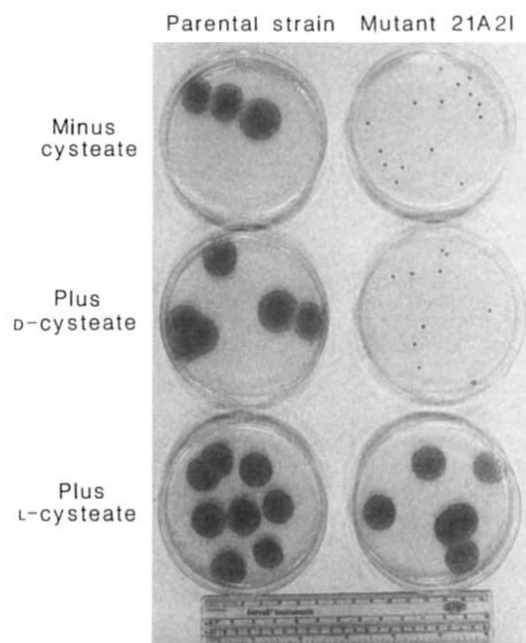


Fig. 2 Colony spreading in *C. johnsonae*. Cells of the parent strain or mutant 21A2I (as indicated) were cloned on LTY medium (0.05% yeast extract, 0.05% Tryptone, solidified with 1.8% agar) containing (where indicated) D- or L-cysteate ($25 \mu\text{M}$). After growing for 84 h at 30°C , the colonies were stained with tetrazolium blue to enhance contrast for photography. A six-inch ruler is shown for scale.

tein⁶; indeed, the compound appears not to be metabolized except by incorporation into capnine. Thus it is not likely that cysteate restores gliding via some metabolic transformation other than conversion to sulphonolipid. These observations support the conclusion that sulphonolipids are necessary for gliding and militate strongly against other interpretations, as for example that 21A2I is a double mutant, or that it is deficient in some cellular product that is independently necessary both for sulphonolipid synthesis and for gliding. Such interpretations require two distinct roles for cysteate in restoring both sulphonolipid content and gliding motility.

Two additional gliding-deficient, sulphonolipid-deficient mutants were studied. Number 36 is phenotypically identical to 21A2I in mutant colony spreading and raft movement. It has normal total lipid sulphur content, over 99% of this in the form of free (non-acylated) capnine and there is no detectable N-acylcapnine (whereas in the parent strain, the converse is true). As expected, provision of L-cysteate does not restore gliding.

Table 1 Properties of mutant 21A2I

Strain	Addition to growth medium	Raft movement ($\mu\text{m min}^{-1}$)	Sulphonolipid content ($\mu\text{mol per g cells, wet wt}$)
Parent	None	25 ± 6	6.2 ± 1.2
Parent	L-cysteate	24 ± 8	10.0 ± 2.6
21A2I	None	None	< 0.1
21A2I	L-cysteate	35 ± 8	9.4 ± 2.6

Where indicated, L-cysteate was present at $25 \mu\text{M}$. Raft movement was measured directly, on colonies grown as described in Fig. 2 legend, using a phase-contrast microscope fitted with a television monitor. Values are the mean ($\pm$ standard deviation) of at least 12 independent determinations; 21A2I showed no raft formation or gliding movement in unsupplemented medium. When cysteate was absent, sulphonolipid content was measured, as described in the text, using $\text{Na}_2^{35}\text{SO}_4$ as the sole sulphur source. Cysteate, when added, was labelled with ^{35}S . Values are the mean ($\pm$ s.d.) of at least six independent determinations.

Evidently, mutant 36 can synthesize cysteate and capnine but is defective in the *N*-acylation of capnine. This suggests that ability to form the *N*-acylated lipid is the specific requirement for gliding, but this conclusion must be regarded as tentative in the absence of supporting evidence. Another mutant, number 15, glides slowly (raft-movement and colony-spreading rates are 28 and 40% of the parent-strain values, respectively). Its lipid sulphur content is 40% of that found in the parent strain; it has traces of both capnine and *N*-acylcapnine, but most of the sulphur is found in abnormal sulphonolipids that do not co-chromatograph with either. Their nature is under investigation.

As has been reported⁸ for many other non-gliding, non-spreading mutants of this bacterium, cells of the four mutants reported here display several behaviours characteristic of wild-type cells when examined microscopically in wet mounts of suspensions on glass slides. These include cell movements described⁹ as spinning and pivoting, latex microsphere (260 nm diameter) movement over cell surfaces, and brief translocational movements of the cells (which, in the case of our mutants and parent strain, rarely exceed a single cell length on glass surfaces). We therefore conclude that sulphonolipids, although required for gliding motility and colony spreading on agar, are not required for the movements of *C. johnsonae* seen in wet mounts.

It seems reasonable to suppose that the cell and cell-surface movements observed in the wet mounts are caused by some subset of the machinery responsible for biologically significant gliding motility (cell translocation over long distances on solid surfaces). At present, however, the evidence⁸⁻¹¹ for this is indirect and the machinery remains undescribed. Our demonstration of the importance of the novel sulphonolipids in gliding motility provides new opportunities for examining the molecular basis of that process in these bacteria and for clarifying the relationship between cell movements in wet mounts and the mass translocation of cells over solid surfaces.

This research was supported by NSF grant PCM8214050.

Received 15 April; accepted 15 September 1986.

1. Reichenbach, H. A. *Rev. Microbiol.* **35**, 339-364 (1981).
2. Burchard, R. P. A. *Rev. Microbiol.* **35**, 497-529 (1981).
3. Godchaux, W. & Leadbetter, E. R. *J. Bact.* **144**, 592-602 (1980).
4. Godchaux, W. & Leadbetter, E. R. *J. biol. Chem.* **259**, 2982-2990 (1984).
5. Godchaux, W. & Leadbetter, E. R. *J. Bact.* **153**, 1238-1246 (1983).
6. Abbanat, D. R., Godchaux, W., Polychroniou, G. & Leadbetter, E. R. *Biochem. biophys. Res. Commun.* **130**, 873-878 (1985).
7. Brady, R. O. & Koval, G. O. *J. biol. Chem.* **233**, 26-31 (1958).
8. Chang, L.-Y. E., Pate, J. L. & Betzig, R. O. *J. Bact.* **159**, 26-35 (1984).
9. Pate, J. L. & Chang, L.-Y. E. *Curr. Microbiol.* **2**, 59-64 (1979).
10. Wolkin, R. H. & Pate, J. L. *J. gen. Microbiol.* **131**, 737-750 (1985).
11. Lapidus, I. R. & Berg, H. C. *J. Bact.* **151**, 384-398 (1982).

Ion channels in human neutrophils activated by a rise in free cytosolic calcium concentration

V. von Tscharner, B. Prod'hom*, M. Baggiolini & H. Reuter*†

Theodor Kocher Institute and * Department of Pharmacology, University of Berne, 3010 Berne, Switzerland

A rapid, transient rise in the free cytosolic Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) is one of the earliest events in neutrophil activation and is assumed to be involved in many of the subsequent cellular reactions¹⁻³. Both Ca^{2+} release from intracellular stores and Ca^{2+} influx from the extracellular space contribute to the rise in $[\text{Ca}^{2+}]_i$ ⁴. In an attempt to assess the relative importance of these pools and the sequences leading to the rise in $[\text{Ca}^{2+}]_i$, we have studied the time course of changes in $[\text{Ca}^{2+}]_i$ after stimulation with *N*-formyl-methionyl-leucyl-phenylalanine (fMLP) or platelet-activating factor (PAF) using the Ca^{2+} indicators quin-2 and fura-2. We observed a time lag of 1-3 s between stimulation and rise in $[\text{Ca}^{2+}]_i$. This lag depends on the agonist concentration but is independent of extracellular Ca^{2+} . Thus Ca^{2+} release from intracellular stores is rate limiting for the rise in $[\text{Ca}^{2+}]_i$. After this, cation channels in the plasma membrane (measured with the patch clamp method⁵) are opened. These non-selective channels, which also pass Ca^{2+} , are activated by the initial rise in $[\text{Ca}^{2+}]_i$, but by neither fMLP nor inositol 1,4,5-trisphosphate (IP_3) directly.

The time courses of change in $[\text{Ca}^{2+}]_i$ were measured in purified human neutrophils loaded with the indicators quin-2 (ref. 6) or fura-2 (ref. 7). Shortly before stimulation the cells were suspended in an isotonic saline solution containing either 1 mM CaCl_2 or 1 mM EGTA with no CaCl_2 . Figure 1 shows the increase in $[\text{Ca}^{2+}]_i$ measured with quin-2 after stimulation of the neutrophils with the calcium ionophore ionomycin or the chemotactic peptide fMLP. A rise in $[\text{Ca}^{2+}]_i$ was observed in both cases in the presence and absence of extracellular Ca^{2+} . The time courses, however, were different: the fluorescence changes induced by ionomycin started in less than 0.2 s, whereas those induced by fMLP were delayed up to 3 s (~1 s with fMLP

10^{-7} M; Fig. 1). The rate of fluorescence increase following either stimulus was about twice as fast when extracellular Ca^{2+} was present, because Ca^{2+} influx from the extracellular space is added to Ca^{2+} release from intracellular stores⁸⁻¹⁰. The experiments with ionomycin show that: (1) $[\text{Ca}^{2+}]_i$ changes mediated by the ionophore occur without a measurable time lag and (2) the time resolution of our testing system (depending on mixing, Ca^{2+} buffering by the indicator and the response time of the apparatus) was well above the range of the lag periods after stimulation with fMLP. Thus the delay in the response to fMLP reflects the time required for stimulus transduction.

We performed further experiments using fura-2 as the indicator and either fMLP or PAF as stimuli. Fura-2 can be used at lower concentrations than quin-2, resulting in less buffering of $[\text{Ca}^{2+}]_i$ because of its higher fluorescence quantum yield⁷. The fura-2 fluorescence changes were comparable to those seen with quin-2. A lag period of similar duration was observed between stimulation with either fMLP or PAF and the increase in $[\text{Ca}^{2+}]_i$. The lag time shortens and the rate of fluorescence change increases markedly with increasing concentrations of fMLP (Fig. 2). Similar families of curves were recorded upon stimulation with PAF (not shown). With either stimulus the duration of the lag was virtually identical with or without extracellular Ca^{2+} (6 experiments).

Thus the rise in $[\text{Ca}^{2+}]_i$ is initiated by a rate-limiting process that does not depend on extracellular Ca^{2+} . But it is clear that Ca^{2+} influx does occur in fMLP- or PAF-stimulated cells (Fig. 1)⁸⁻¹⁰. What controls the Ca^{2+} permeability of the plasma membrane? We have performed patch clamp experiments⁵ and have measured ion channel activity in the neutrophil plasma membrane directly. Channels with similar properties were recorded in both cell-attached and isolated, inside-out membrane patches. In resting cells the occurrence of channel openings was extremely rare and the opening probability was not increased in the presence of fMLP (10^{-8} to 10^{-6} M) in the pipette solution. This strongly suggests that the channels are not directly coupled to the agonist receptor. If fMLP was added to the bathing medium with Ca^{2+} concentrations of 1-2 mM, about half (13/24) of the cells responded with increased channel openings. If cells were loaded with fura-2 in the absence of external Ca^{2+} (ref. 11) $[\text{Ca}^{2+}]_i$ was reduced to about 10^{-9} M and fMLP did not induce channel openings. Extensive single channel activity was observed in inside-out membrane patches, provided the Ca^{2+} concentration in the bathing solution (which is in contact with

† To whom correspondence should be addressed.

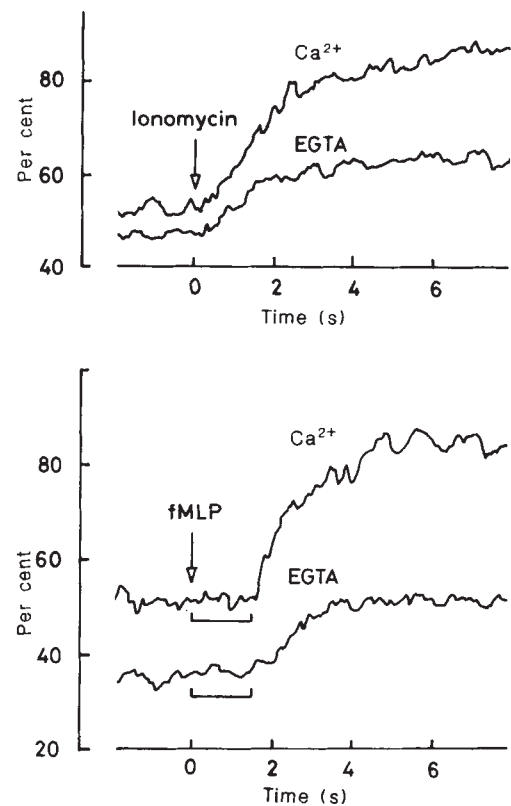
Fig. 1 Quin-2 fluorescence curves in human neutrophils following addition (arrow) of ionomycin (2×10^{-6} M) or fMLP (10^{-7} M). The upper traces (Ca^{2+}) were obtained in the presence of 1 mM CaCl_2 in the medium, and the lower traces (EGTA) in a calcium-free medium with 1 mM EGTA added 5 min before stimulation. The lag times between fMLP addition and onset of quin-2 fluorescence increase are indicated by the bars underneath the tracings.

Methods. Neutrophils were obtained from buffy coats of donor blood stored for up to 24 h at $4-10^\circ\text{C}$ (Swiss Red Cross Laboratory, Berne, Switzerland) and purified by centrifugation through a discontinuous Lympho-paque gradient²¹. The cells collected at the boundary between the two density layers (95–98% neutrophils) were washed three times by centrifugation (160g for 5 min at room temperature) in a saline medium containing 136 mM NaCl, 4.8 mM KCl, 1.2 mM KH_2PO_4 , 5 mM glucose, buffered at pH 7.4 with 20 mM *N*-2-hydroxyethylpiperazine-*N*-2-ethanesulphonate (HEPES/NaOH). The final suspension was adjusted to 4×10^6 cells per ml and kept at room temperature until use. Loading with the fluorescent indicator: Neutrophils (4×10^6 cells per ml) were incubated for 30 min at 37°C in the presence of quin-2 acetoxy-methyl ester (0.4 nmol per 10^6 cells) dispensed from a 2 mM stock solution in dimethyl sulfoxide. The process was followed fluorimetrically and was considered to be complete when a maximum was observed in the 510/465 (nm/nm) fluorescence intensity ratio (510 nm is just above the fluorescence peak; 465 nm is an arbitrary reference point that changes very little as loading proceeds). A filter fluorimeter was specially constructed for high time resolution measurements. The excitation wavelength was selected from a 500 W Osram Hg-lamp with a band filter of 334 ± 10 nm. A 470 nm cut-off filter was inserted on the emission side allowing the measurement of the whole emission band with a photomultiplier (EMI 9798QB) connected to an amplifier with a response time of 30 ms. Broad band emission measurement and the indicated amplifier's response time resulted in a reasonably high signal-to-noise ratio. Loaded neutrophils were stimulated at 37°C in disposable polystyrol cuvettes (8 mm inner diameter) under magnetic stirring. The agonist was injected through a light-tight opening in the lid of the cuvette holder with a Hamilton syringe. Injection was timed by a triggering signal to the recorder. Intracellular calibration of the quin-2 signal was used to determine the quin-2 saturation fraction q . Ionomycin (2×10^{-6} M) was added to the cells in the presence of 1 mM extracellular Ca^{2+} to bring about complete saturation of intracellular quin-2 and the fluorescence was taken as F_{max} . All quin-2 fluorescence was then quenched by addition of 10 mM MnCl_2 and the background fluorescence was taken as F_0 . In agreement with earlier work²² the fraction of total quin-2 present as calcium complex q is then calculated according to the relations:

$$q = \frac{(F - F_{\text{min}})}{(F_{\text{max}} - F_{\text{min}})}$$

where

$$F_{\text{min}} = F_{\text{max}} - \frac{5}{6}(F_{\text{max}} - F_0)$$



the inner surface of the patch) was higher than 10^{-7} M. We have also measured the activity of these channels in saponin treated cells. We exposed whole cells to low (0.005%) saponin concentrations for ~10 min after establishing a gigaohm seal between pipette tip and membrane¹². Saponin makes membranes permeable to small molecules. If Ca^{2+} was present in the external medium (10^{-5} – 10^{-3} M) we observed activation of ion channels identical to those seen in inside-out patches, but if the Ca^{2+} concentration was lowered to 10^{-8} M the activity disappeared. Two types of single channel currents could be identified (Fig. 3). Although we have no direct evidence that the two current levels in Fig. 3 result from openings of two distinct populations of ion channels, in experiments with excised inside-out patches the larger currents disappeared more rapidly (within a few minutes) than the smaller ones, transitions between current levels were not strictly coupled and in many patches only one type of current could be seen. Therefore, we suggest that the currents flow through two separate types of ion channels.

With identical solutions (130 mM KCl or K-aspartate) on both sides of an inside-out membrane patch the current-voltage relationships were roughly linear with slope conductances of 18–25 pS and 4–6 pS at about 24°C , respectively. Both channels discriminate only poorly between Na^+ , K^+ and Ca^{2+} . This was assessed by measuring the reversal potential of single channel currents with two different ion species on either side of excised inside-out patches of membrane (Fig. 4a,b). With 90 mM CaCl_2 solution in the pipette and 130 mM KCl or K-aspartate solution

in the bath, the reversal for both channel currents was near 0 mV (Fig. 4d), indicating that both channels were permeable to Ca^{2+} and K^+ . The same was true when a 130 mM NaCl solution replaced the K-aspartate solution in the bath or the CaCl_2 solution in the pipette. Outward currents through the channels at positive membrane potentials were usually larger than inward currents (Fig. 4d). Channel openings were seen more often at positive than at negative membrane potentials, implying some voltage dependence of channel gating. However, single channel currents could only be measured if the Ca^{2+} concentration in the bath was raised to above 10^{-7} M (Figs 3c and 4c). Therefore, a rise in $[\text{Ca}^{2+}]_i$ is an essential step in the gating of both channels. Open times ranged from a few milliseconds up to several seconds. The smaller conductance channel was more likely to stay open for many seconds. IP_3 , which has been shown to release Ca^{2+} from intracellular stores in neutrophils¹³, does not by itself induce ion channel activity when added to the internal surface of the membrane (Fig. 3b).

Our findings strongly suggest that in neutrophils the first step leading to a change in $[\text{Ca}^{2+}]_i$ is the release of Ca^{2+} from intracellular stores. This Ca^{2+} release seems to be induced by IP_3 (ref. 13) formed after agonist-receptor occupancy¹⁴. The increase in $[\text{Ca}^{2+}]_i$ due to Ca^{2+} release causes transient openings of Ca^{2+} -gated ion channels in the membrane through which more Ca^{2+} ions flow into the cytoplasm. Over a time period of 1–2 min the fluorescence signals decrease again (not shown) indicating a reduction in the free cytoplasmic Ca^{2+} concentra-

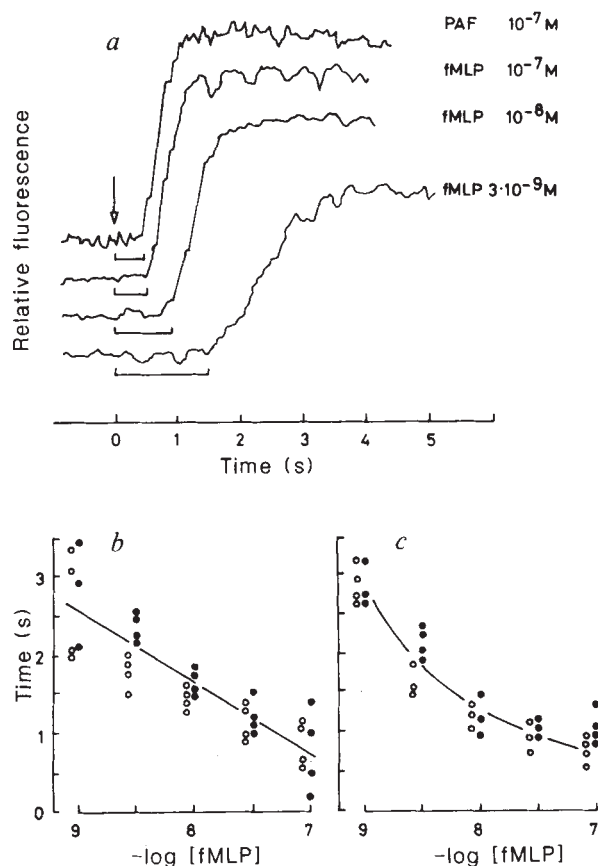


Fig. 2 *a*, Fura-2 fluorescence increase in human neutrophils stimulated with PAF or fMLP. The lag times between agonist addition (arrow) and the onset of fluorescence increase are indicated by the bars under the single tracings. Average $[Ca^{2+}]_i$ prior to stimulation with agonists was 1.3×10^{-7} M. *b*, Plot of the lag time between fMLP addition and onset of fluorescence increase against fMLP concentration. Single values obtained with the same cell preparation in the presence (●) and absence (○) of extracellular calcium are shown with the calculated regression line. No statistical difference was observed between the regression lines through the values obtained with or without extracellular calcium. *c*, Time for an increase in fluorescence to 90% of maximum against concentration of fMLP. Data from the same series of experiments as *b*.

Methods. The neutrophils were prepared as described (Fig. 1) and loaded with fura-2 (0.1 nmol per 10^6 cells) using the same protocol as for quin-2. Fluorescence measurements were made with the instrument described in Fig. 1 legend. Excitation was at 334 ± 10 nm and emission was recorded through a band-pass filter of 500 ± 10 nm. Since lag time measurements do not require it, absolute calibration of the fura-2 signal was omitted to allow frequent repeats of the assay with single batches of cells. A fluorescence decrease due to degranulation as seen in mast cells²³ was not observed.

tion. In intact cells we found spontaneous inactivation of the cation channels which may contribute to this fall in $[Ca^{2+}]_i$. Since the opening of the poorly selective cation channels is not rate limiting for the initial rise in $[Ca^{2+}]_i$, the observed lag period is independent of the external Ca^{2+} concentration. Influx of Na^+ in addition to Ca^{2+} ions through these channels accounts for the depolarization associated with neutrophil stimulation¹⁵. We found no evidence for voltage-dependent Ca^{2+} channels as in neoplastic B lymphocytes¹⁶. The channels described in this paper also differ from other Ca^{2+} -activated non-selective cation channels^{17,18} in being highly permeable to Ca^{2+} ions. Although the exact role of Ca^{2+} in other neutrophil reactions (for example, superoxide anion production, degranulation, aggregation) is still

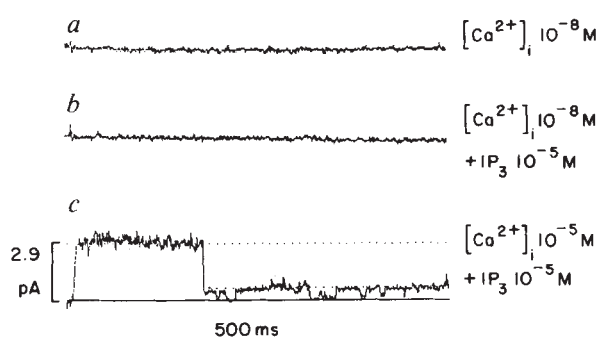


Fig. 3 Ion channel activation by $[Ca^{2+}]_i$ in an inside-out patch of neutrophil membrane. *a*, 10^{-8} M $[Ca^{2+}]_i$; *b*, 10^{-8} M $[Ca^{2+}]_i$ + 10^{-5} M IP_3 ; *c*, 10^{-5} M $[Ca^{2+}]_i$ + IP_3 . No channel openings were seen in *a* or *b*, but two outward current levels resulting from openings of ion channels can be seen in *c*. Presence of IP_3 was not required for activation of the channels. The membrane potential of the patch was held at -100 mV and currents were recorded during voltage steps of 500 ms duration to $+100$ mV.

Methods. Single channel currents were recorded in an excised, inside-out patch of membranes from human neutrophils using the patch clamp method⁵. The bath solution (23–25 °C) contained 130 mM K-aspartate, 20 mM HEPES buffer (pH 7.4) and 10^{-8} M $[Ca^{2+}]_i$ (0.32 mM $CaCl_2$ + 1 mM EGTA; apparent stability constant $10^{7.5}$ M⁻¹), or 10^{-5} M (unbuffered) free $[Ca^{2+}]_i$; the solution in the recording pipette contained 90 mM $CaCl_2$ and 10 mM HEPES buffer (pH 7.4). Digitized records (1 ms per point) were analysed by a computer (PDP 11/04). Capacity transients and leakage currents during voltage steps have been subtracted. Heavy lines in the records indicate the closed state of the channel while dotted lines show the average current levels when the channels are open.

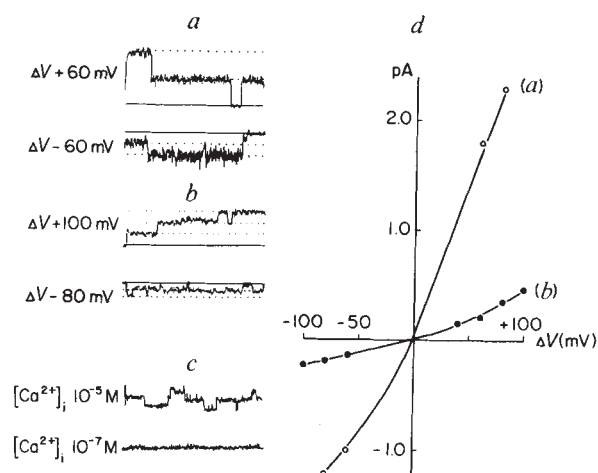


Fig. 4 Examples of current traces (*a–c*) and corresponding current-voltage relationships (*d*) of two types of ion channels in excised inside-out membrane patches from human neutrophils. Bath and pipette solutions were the same as in Fig. 3. Outward currents through open channels (upward deflections from heavy base line) were carried by K^+ ions, while inward currents (downward deflections) were carried by Ca^{2+} ions (*a*, *b*). Channel openings were observed only when the $[Ca^{2+}]_i$ at the inner surface of the membrane patch was raised above 10^{-7} M (*c*). Time scales are 0.5 s in *a* and *b* and 10 s in *c*. Current amplitudes during single channel openings at different membrane potentials have been plotted in *d*. ○, Current-voltage relationship from single-channel currents in *a*; ●, current-voltage relationship from single-channel currents in *b*. Note that outward currents at positive pulse potentials (ΔV) are larger than inward currents at negative potentials. The reversal potential is near 0 mV, indicating similar permeabilities for K^+ and Ca^{2+} through the channel.

unclear¹⁹, the rise in $[Ca^{2+}]_i$ appears to be an important intermediary step in neutrophil activation^{15,20}.

We thank the Swiss National Science Foundation (grants 3.353-0.82 to M.B. and 3.059-0.84 to H.R.) for generous support and Markus Gerber for technical assistance.

Received 21 April; accepted 24 September 1986.

1. Pozzan, T., Lew, D. P., Wollheim, C. B. & Tsien, R. Y. *Science* **221**, 1413-1415 (1983).
2. White, J. R., Naccache, P. H., Molski, T. F. P., Borgeat, P. & Sha'afi, R. I. *Biochem. biophys. Res. Commun.* **113**, 44-50 (1983).
3. Gennaro, R., Pozzan, T. & Romeo, D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1416-1420 (1984).
4. Korchak, H. M. et al. *J. biol. Chem.* **259**, 4076-4082 (1984).
5. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
6. Tsien, R. Y. *Biochemistry* **19**, 2396-2404 (1980).

7. Grynkiewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440-3450 (1985).
8. Sklar, L. A. et al. *J. biol. Chem.* **260**, 11461-11467 (1985).
9. Sklar, L. A. & Oades, Z. G. *J. biol. Chem.* **260**, 11468-11475 (1985).
10. von Tscherner, V., Derenleau, D. A. & Baggiolini, M. *J. biol. Chem.* **261**, 10163-10168 (1986).
11. Di Virgilio, F., Lew, D. P. & Pozzan, T. *Nature* **310**, 691-693 (1984).
12. Petersen, O. H. & Petersen, C. C. H. *News physiol. Sci.* **1**, 5-8 (1986).
13. Spät, A., Bradford, P. G., McKinney, J. S., Rubin, R. P. & Putney, J. W. Jr *Nature* **319**, 514-516 (1986).
14. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
15. Korchak, H. M., Vienne, K., Rutherford, L. E. & Weissmann, G. *Fedn Proc.* **43**, 2749-2754 (1984).
16. Fukushima, Y. & Hagiwara, S. *J. Physiol., Lond.* **358**, 255-284 (1985).
17. Colquhoun, D., Neher, E., Reuter, H. & Stevens, C. F. *Nature* **294**, 752-754 (1981).
18. Yellen, G. *Nature* **296**, 357-359 (1982).
19. Barrowman, M. M., Cockcroft, S. & Gomperts, B. D. *Nature* **319**, 504-507 (1986).
20. Sawyer, D. W., Sullivan, J. A. & Mandell, G. L. *Science* **230**, 663-666 (1985).
21. Bøyum, A. *Scand. J. Clin. Invest.* **21** (Suppl. 97), 77-79 (1986).
22. Heskest, T. R., Smith, G. A., Moore, J. P., Taylor, M. V. & Metcalfe, J. C. *J. biol. Chem.* **258**, 4876-4882 (1983).
23. Almers, W. & Neher, E. *FEBS Lett.* **192**, 13-18 (1985).

A complementary DNA clone for a macrophage-lymphocyte Fc receptor

Victoria A. Lewis, Terry Koch,
Helen Plutner & Ira Mellman*

Department of Cell Biology, Yale University School of Medicine,
333 Cedar Street, PO Box 3333, New Haven,
Connecticut 06510, USA

Macrophages, granulocytes and many lymphocytes express or secrete receptors for the Fc domain of immunoglobulins (Ig)¹. These Fc receptors (FcRs) are heterogeneous and can be distinguished on the basis of their cellular distribution and specificities for different immunoglobulin isotypes. Although their functions are not completely understood, FcRs are known to be involved in triggering various effector cell functions and in regulating differentiation and development of B-cells¹⁻⁴. One of the best characterized is the mouse macrophage-lymphocyte receptor for IgG₁ and IgG_{2b} (ref. 5). On macrophages, this FcR mediates the endocytosis of antibody-antigen complexes via coated pits and coated vesicles^{6,7}, the phagocytosis of Ig-coated particles⁸, and the release of various inflammatory and cytotoxic agents¹. It is possible that the receptor possesses an intrinsic ligand-activated ion channel activity responsible for some of these functions^{9,10}. The IgG₁/IgG_{2b} FcR has been isolated and shown to be a transmembrane glycoprotein of relative molecular mass (M_r) 47,000-60,000 (47-60 K) containing four N-linked oligosaccharide chains and a large (>10K) cytoplasmic domain¹¹. It is also immunologically indistinguishable from the murine Ly-17 alloantigen which, in turn, is tightly linked to the *Mls* lymphocyte activation locus¹²⁻¹⁴. Here we describe the isolation and characterisation of a complementary DNA clone encoding the whole of the IgG₁/IgG_{2b} FcR expressed by the mouse macrophage-like cell line P388D₁. The receptor is a member of the immunoglobulin superfamily and, like Ly-17, maps to the distal portion of chromosome 1. cDNA probes detect one or two mRNA species in FcR⁺ macrophage and B-cell lines, but not in FcR⁻ cells or a receptor-deficient variant derived from a FcR⁺ B-cell line¹⁵. Finally, DNA hybridization analysis indicates the receptor gene is partially deleted or rearranged in the FcR⁻ variant.

FcR cDNAs were isolated from a P388D₁ λgt11 library (obtained from A. Ezekowitz, Harvard University) using both antibody and oligonucleotide probes. Antibody screening was performed¹⁶ using a monospecific rabbit antiserum raised against sodium dodecyl sulphate (SDS)-denatured FcR purified from the J774 mouse macrophage cell line¹¹. A single positive phage was plaque-purified and found to contain a 580 base pair (bp) insert (pFcRab). For oligonucleotide screening, a double stranded 42 bp probe corresponding to residues 2-15 of the FcR's chemically determined N-terminal amino-acid sequence was made (Fig. 1a). Two plaques (pFcR11 and pFcR13) were

purified, subcloned and found to contain 0.8 and 1.5 kilobase (kb) inserts respectively. The clone isolated by antibody screening cross-hybridized with those obtained by oligonucleotide screening (not shown). Their specificity for the FcR was confirmed by nucleotide sequence analysis of the 5' ends of pFcR11 and pFcR13. Both contained sequences corresponding to the N-terminal amino acid-sequence (Fig. 1b).

Figure 1c shows the complete nucleotide sequence compiled from the three inserts, and the amino acid sequence of the FcR thereby deduced. The largest clone, pFcR13, includes the receptor's entire coding region as well as 3' untranslated sequences containing a consensus poly(A)⁺ addition sequence and a poly(A)⁺ tract. pFcR11 corresponds to the 5' half of the 1.5 kb pFcR13 clone, whereas pFcRab corresponds to the 3' half: they overlap with each other by only 27 nucleotides. From the deduced amino acid sequence, the FcR contains an N-terminal hydrophobic signal peptide. However, some ambiguity remains regarding the identity of the actual initiator AUG codon as the putative signal sequence contains three methionine residues (at positions -13, -29 and -39) any of which could begin a signal sequence. We believe that Met-13 is the best candidate because the putative signal it begins contains the hydrophobic and cleavage domains typical of shorter signal peptides¹⁷. Although Met-29 lies in a nucleotide sequence most closely (but not precisely) corresponding to a consensus start site for translation¹⁸, it implies a signal peptide beginning with an atypically long stretch of polar residues (G. von Heijne, personal communication).

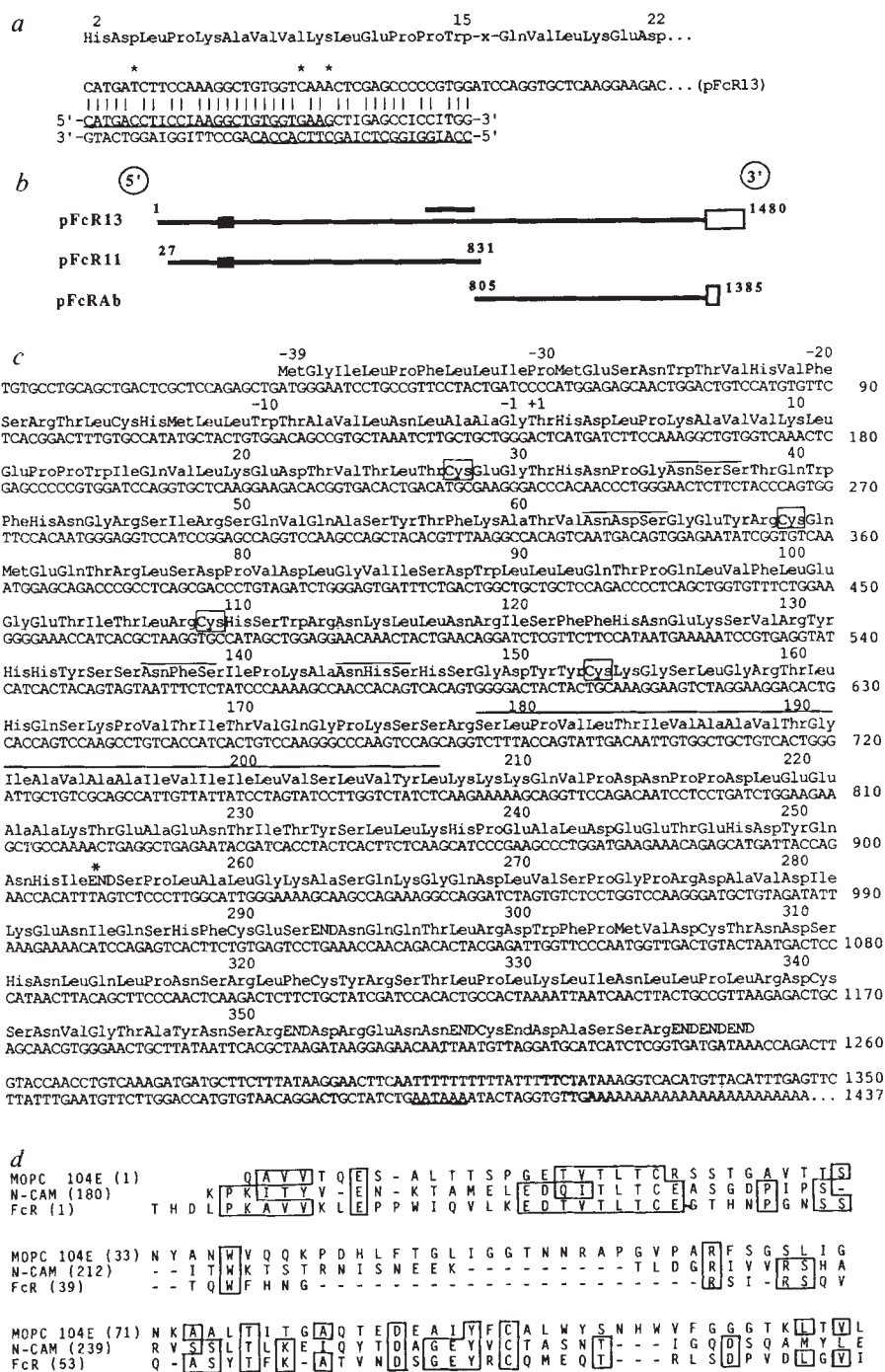
The amino acid sequence of the mature FcR starts with a Thr and contains four potential N-linked glycosylation sites (Asn-X-Ser/Thr, X ≠ Pro; at positions 36, 63, 137 and 144), a single hydrophobic membrane-spanning segment followed by a charged stop-transfer sequence, and a large hydrophilic cytoplasmic domain. These results agree well with the results of our previous biochemical investigations which used endoglycosidase digestion to identify four N-linked oligosaccharides and protease digestion of ³⁵S-methionine-labelled microsomes to locate a cytoplasmic segment possibly >10 kD in length¹¹. While the length of the cytoplasmic domains predicted from the position of the first stop codon is only 47 residues (6.7 K), the size determined from SDS polyacrylamide gel electrophoresis data could easily be an overestimation, as found previously¹⁹. Nevertheless it is interesting that the first stop codon is followed by two long open reading frames of 32 and 65 amino acids prior to the occurrence of multiple stop sites. Although the digestion experiments also showed that ~15 K of the receptor was resistant to proteolysis from either side of the membrane, the nucleotide sequence reveals that this protection was not due to the presence of multiple membrane-spanning domains. As the FcR contains only a single hydrophobic domain it is probable that its putative ion channel activity^{9,10} could occur only by receptor aggregation. However, the FcR's predicted membrane spanning segment of 29 amino acids is somewhat longer than those found in most membrane proteins.

* To whom correspondence should be addressed.

Fig. 1 IgG₁/IgG_{2b} Fc receptor nucleotide sequence and predicted amino acid sequence. *a*, Nucleotide sequence of synthetic oligonucleotide probe based on N-terminal amino-acid sequence of the intact FcR protein isolated from J774 cells. Assignment of residue 1 (Thr) was ambiguous and therefore omitted, as was residue 16 (Ile). Overlapping 27- and 24-mers (underlined) corresponding to amino acids 2-15 from coding and non-coding strands were chemically synthesized. Five deoxynosine residues were inserted at positions of highest ambiguity. Asterisks denote three positions of mismatch between the sequence of the synthetic probe and the isolated cDNA clone. *b*, Schematic representation of cDNA clones representing the complete FcR coding sequence. pFcRab and pFcR11 correspond to the 3' and 5' halves of the coding sequence respectively and overlap by 27 nucleotides just 3' to sequences coding for the putative transmembrane domain (overlined). pFcR13 completely contains both pFcRab and pFcR11. Black boxes, sequences corresponding to the synthetic oligonucleotide probe. Open boxes, poly(A)⁺ tracts. *c*, Nucleotide and predicted amino-acid sequence of pFcR13. Nucleotides are numbered in the 5' to 3' direction starting with the first nucleotide in the codon of the N-terminal residue in the mature protein. Nucleotides 5' to base 1 have negative numbers and include residues encoding a putative signal peptide and a 5' untranslated region. The four glycosylation sites are lightly overlined, cysteine residues involved in forming Ig-like domains are boxed, and the consensus polyadenylation signal is underlined. The single putative transmembrane domain is indicated by the heavy overline, and is followed by a hydrophilic stop-transfer sequence of three lysine residues. *d*, Amino-acid sequence alignment of FcR, N-CAM, and a mouse λ -light chain (MOPC 104E) Ig-like domains. Common residues are boxed, and gaps inserted by dashes to achieve maximum alignment. The amino-acid residue that begins each line is given in parentheses. FcR sequence corresponds to the first Ig domain; sequence of the MOPC 104E variable region is given. Similar alignments can be obtained using rabbit polymeric IgR, murine Thy-1, Ia antigen β -chain, or Lyt-2 sequences. We thank Alan F. Williams (MRC, Oxford) for assistance with this analysis.

Methods. *a*, FcR was purified to homogeneity by immunoaffinity chromatography using an anti-FcR monoclonal antibody, 5F2a¹¹, coupled to a 0.5 ml epoxy-activated, silicon-based HPLC matrix provided by Beckman Instruments (final concentration 30 mg IgG per ml resin). The detergent phase of a TX-114 lysate was prepared from 5 ml packed J774 cells (as previously described for immunoprecipitation¹¹), filtered and passed once over the column at 0.2 ml min⁻¹ at 4°C. Following extensive washing with Tris-buffered saline, bound material was eluted with 0.5 M propionic acid containing 25 mM β-D-octylglucoside, immediately neutralized and dialysed exhaustively against glass-distilled water. Final yield was 39 μg of FcR per 2.5 × 10⁹ cells. N-terminal sequence analysis was performed on a modified Beckman 890C spinning cup sequencer and allowed the unambiguous assignment of 20 contiguous amino acids with gaps at positions 1 and 16. *b*, A P388D₁ Agt-11 cDNA library was screened using a monospecific polyclonal anti-FcR antiserum followed by ¹²⁵I-labelled protein A essentially as described¹⁶. A single clone was identified by screening a sublibrary using a double-stranded, high-specific-activity oligonucleotide probe as described¹⁶. The cDNA was subcloned into pUC9 or Bluescript plasmid vectors. *c*, DNA sequencing was performed on both strands of the cDNA using the Maxam and Gilbert²⁸ method (on only one strand) by the method of Maxam and Gilbert²⁸.

as MOPC 104E, and N-CAM²¹ (Fig. 1d). Thus the patterns around the Cys residues seem closest to V and V-like domains but the length of sequence between the Cys residues is far too short to account for a V-like fold. If β -strands are assigned on the basis of sequence similarities then it seems possible that the FcR domains may be folded in the C-domain pattern (A. F. Williams, personal communication). Computer searches of nucleic acid and protein sequence databases (GenBank, Staden) indicated further similarities between the FcR and other immunoglobulin-like molecules including class II MHC antigen β -chain, Lyt-2 antigen and murine Thy-1 (which, like N-CAM,



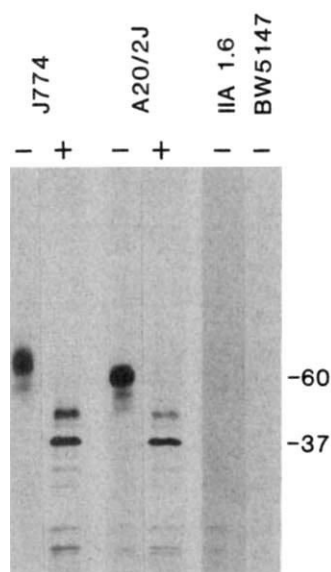


Fig. 2 Immunoprecipitation of ^{35}S -FcR from macrophage, B- and T-lymphoma cell lines. J774, A20/2J, IIA-1.6 and BW5147 cells were metabolically labelled with ^{35}S -methionine, lysed using Triton X-114 and immunoprecipitated with anti-FcR antiserum as described previously¹¹. Radiolabelled bands corresponding to the mature receptor were detected in lysates of J774 cells (64K) and A20/2J cells (60K). Endoglycosidase F treatment (+ lanes) of immunoprecipitated ^{35}S -FcR gave identical 37K peptides (arrow), and several minor bands resulting from incomplete digestion or proteolytic degradation¹¹. Both the T-lymphoma BW5147 and the selected variant of A20/2J, IIA-1.6, failed to synthesize immunoprecipitable FcR. Molecular weights in kilodaltons.

is expressed in neural tissues)²¹.

As mentioned above, the $\text{IgG}_1/\text{IgG}_{2b}$ FcR has been shown to be immunologically related to the Ly-17 alloantigen, a polymorphic locus which maps to mouse chromosome 1 (refs 12, 13). Thus, it was of interest to determine the chromosomal localization of the gene recognized by FcR cDNA. Using labelled pFcRab, we probed DNA from a chromosome panel derived from CHO $\times$ mouse fibroblast somatic cell hybrids (provided by D. Pravtcheva and F. Ruddle)²². Conditions were chosen for an *Eco*RI digestion such that pFcRab hybridized with a single murine band that was easily differentiated from its CHO homologue (Fig. 4). Analysis of DNA from 9 hybrid cell lines consistently assigned the FcR gene to mouse chromosome 1 (not shown). We also probed DNA from a second CHO $\times$ mouse panel (provided by C. Kozak) that included cell lines characterized by having undergone translocations involving only portions of mouse chromosomes (ref. 23). These results indicate that the FcR structural gene, and therefore the Ly-17 structural gene, is uniquely localized to the region of chromosome 1 distal to the *Pep-3* locus¹⁴.

To confirm that the cDNAs detected mRNA only in FcR⁺ cells, and to investigate the relationship between receptor species expressed in different cell types, hybridization analysis was performed using poly(A)⁺ RNA isolated from FcR⁺ or FcR⁻ cells. Because the presence or absence of FcR is usually determined only on the basis of surface expression¹, we first analysed cell lines by immunoprecipitation after labelling with ^{35}S -methionine. Using a rabbit anti-FcR antiserum, labelled receptor was precipitated from two macrophage cell lines (J774 and P388D₁, only the former is shown) and one B-cell line (A20/2J) (Fig. 2). Although the apparent relative molecular masses of the macrophage and B-cell products were different (64K against 60K), the difference was due to glycosylation heterogeneity, as Endo F treatment of the two immunoprecipitates yielded a 37K deglycosylated peptide. Two cell lines, T-cell line BW5147 and

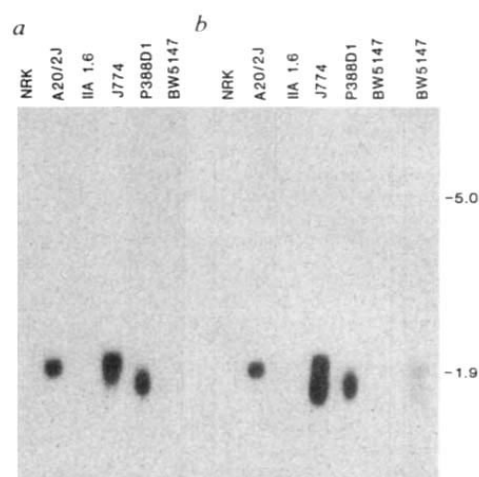


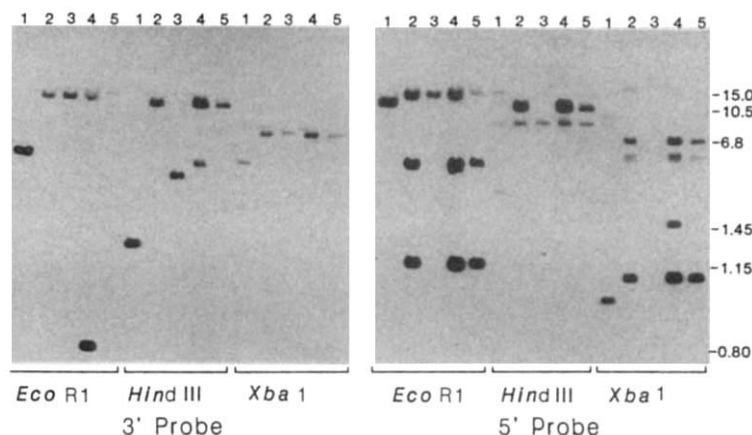
Fig. 3 Northern blot analysis of poly(A)⁺ RNA isolated from FcR-positive and -negative cell lines. RNAs were prepared by guanidine-HCl lysis and phenol/chloroform extraction. Poly(A)⁺ RNA was selected by oligo-dT cellulose chromatography, resolved on 1.0% agarose-formaldehyde gels, and transferred to Zetabind nylon membranes. RNAs were hybridized with the 3' encoding clone, pFcRab, *a*, stripped at 95 °C with 0.1 $\times$ SSC/0.1% SDS, and re-hybridized using the 5' encoding clone, pFcR11, *b*. Purified inserts were labelled by nick-translation to a specific activity of 0.5×10^9 c.p.m. per μg . Hybridization and washings (0.2 $\times$ SSC/0.1% SDS at 50 °C) were carried out essentially as described²⁹. Cell lines used included rat fibroblasts (NRK, FcR⁻), A20/2J B-cells (FcR⁺), the FcR⁻ A20 variant IIA-1.6, the macrophage-like cell lines J774 and P388D₁ (both FcR⁺), and the T-cell lymphoma BW5147 (FcR⁻). Autoradiographs represent 5 h exposures, except for BW5147 which was exposed for 5 h and 48 h (panel B). The 3' probe pFcRab hybridized with single bands only in FcR⁺ cell lines (1.9 kb in A20/2J and J774, 1.7 kb in P388D₁) (panel A). The 5' probe pFcR11 recognized an additional lower molecular weight band (1.6 kb) in J774 cells and in BW5147 cells after long exposures (using the 3' probe, no hybridization was detected after similarly long exposures). Sizes of murine 28S (5.0 kb) and 18S (1.9 kb) ribosomal RNAs are indicated.

a variant of the A20/2J B-cell line, IIA-1.6 (obtained from B. Jones), selected *in vitro* for the loss of surface FcR¹⁵ were shown not to synthesize immunoprecipitable FcR (Fig. 2). Expression of other surface antigens (such as surface immunoglobulin and Ia) in the IIA-1.6 variant, however, is the same as its normal parent A20/2J¹⁵.

Northern blot analysis using pFcRab as a probe only detected single transcripts in mRNA from the FcR⁺ cell lines (Fig. 3*a*). The J774 macrophage and A20 B-cell lines contained mRNAs of similar sizes (1.9 kb) whereas the P388D₁ transcript appeared slightly smaller (1.7 kb). Somewhat different results were obtained using pFcR11 as a probe. Although pFcR11 hybridized with all the transcripts detected by pFcRab, an additional band (1.6 kb) was detected in J774 mRNA (Fig. 3*b*). Because pFcR11 corresponds to the 5' half of the FcR cDNA and pFcRab corresponds to the 3' half (Fig. 1*b*), these results suggest that some cell lines (for example, J774) may make two related FcR species which are more homologous at their N-terminal than at their C-terminal regions. We are now determining whether multiple FcR isoforms are indeed expressed.

Although the 5' probe (pFcR11) detected small amounts of the 1.9 and 1.6 kb transcripts in BW5147 cells, neither the 3' probe (pFcRab) nor the 5' probe hybridized with poly(A)⁺ RNA from fibroblasts or from the FcR⁻ A20/2J variant, IIA-1.6 (Fig. 3). To investigate the mechanism underlying the loss of FcR-specific mRNA from the variant cell line, a preliminary Southern blot analysis was performed. Using the 3'-end probe pFcRab, single bands were detected in J774 and A20/2J DNA digested with *Eco*RI (15 kb), *Hind*III (10.5 kb), or *Xba*I (6.8 kb)

Fig. 4 Southern blot analysis of FcR-positive and -negative celllines. DNA isolated from: track 1, CHO cells (FcR⁻); track 2, A20/2J; track 3, IIA-1.6; track 4, BW5147; and track 5, J774 cells was digested to completion with *Eco*RI, *Hind*III or *Xba*I, separated on a 0.8% agarose gel (10 µg per track) and transferred to nylon (Zetabind) membrane. The blots were then hybridized consecutively with the 3' probe pFcRab (left panel) or the 5' probe pFcR11 (right panel) essentially as described²⁹. Isolated inserts were labelled by nick-translation, filters washed with 0.2×SSC, 0.1% SDS at 55 °C, and the autoradiographs exposed overnight. Sizes of the major restriction fragments are shown in kb.



(Fig. 4, left panel). In contrast, *Hind*III digestion of IIA-1.6 DNA produced a single hybridizable fragment (4 kb), much smaller than that found in the A20/2J parent (10.5 kb). *Hind*III and *Eco*RI digests of the FcR⁺ BW5147 DNA produced additional smaller bands (5.0 kb and 0.8 kb, respectively) not seen in FcR⁺ cell lines. When the same blots were hybridized with the 5'-end probe pFcR11, 1–3 extra bands were detected in all cell lines except the receptor-negative variant IIA-1.6 (Fig. 4, right panel). In IIA-1.6 DNA, the 5' probe even failed to detect the major 4 kb *Hind*III and 6.8 kb *Xba*I fragments seen with the 3' probe (Fig. 4, right panel). As before, pFcR11 also hybridized with a unique low molecular weight band in BW5147 DNA. These data strongly suggest that the failure of the IIA-1.6 variant, and possibly BW5147, to express hybridizable FcR mRNA is attributable to deletions or rearrangements of the FcR gene.

Because the IgG₁/IgG_{2b} receptor studied here is only one of a large family of isotype- and cell type-specific FcRs¹, the existence of multiple FcR-related genes is not surprising. The availability of FcR cDNA clones and cell lines containing alterations in the receptor gene will now permit analysis of the distribution, structure and activities of these receptors, and the definitive establishment of their relationship to functionally-defined genetic loci of immunological importance, such as *Mls*^{12–14,24}.

We thank Adrian Hayday, Elisabetta Ullu, Chris Tschudi, Arthur Horwich, Saumyen Sarkar, Johanna Griffin, Christine Howe, Susan Froschauer and Heini Miettinen for their advice and assistance; Gary Davis and David Speicher for protein sequencing; Alan Williams and Gunnar von Heijne for help with sequence analysis; Christine Kozak, Dimi Pratcheva and Frank Ruddle for providing DNA from their chromosome panels; Alan Ezekowitz and Brigitte Huber for providing cDNA libraries; Barry Jones for providing the IIA-1.6 cell line and Pam Ossorio for photographic work. Supported by NIH grants and a Presidential Research Award from The Leukemia Society of America, Inc.

Note added in proof: During preparation of this manuscript, we learned that Ravage *et al.* (*Science*, in the press) have also cloned and sequenced FcR specific cDNAs. The sequence they obtained from a clone designated β -2 (isolated from a J774 macrophage cell line cDNA library) is identical to that described here (P388D1-derived cDNA). Although identical, the two sequences arise from mouse strains representing different *Mls* alleles (BALB/c: *Mls* b and DBA/2: *Mls* a, respectively). Therefore it is clear that the FcR cloned here, although immunologically identical to the Ly-17 alloantigen, cannot be the product of the *Mls* locus.

Received 14 August; accepted 24 September 1986.

1. Unkeless, J. C., Fleit, H. & Mellman, I. *Adv. Immun.* **31**, 247–270 (1981).
2. Phillips, N. E. & Parker, D. C. *J. Immun.* **132**, 627–632 (1984).
3. Bijsterbosch, M. K. & Klaus, G. G. B. *J. exp. Med.* **162**, 1825–1836 (1985).
4. Daeron, M., Yodoi, J., Neaupoit-Sautes, C., Moncuit, J. and Fridman, W. H. *Eur. J. Immun.* **15**, 662–667 (1985).
5. Mellman, I. & Unkeless, J. C. *J. exp. Med.* **152**, 1048–1069 (1980).
6. Ukkonen, P., Lewis, V. A., Marsh, M., Helenius, A. & Mellman, I. *J. exp. Med.* **163**, 952–971 (1986).
7. Mellman, I. & Plutner, H. *J. Cell Biol.* **98**, 1170–1176 (1984).
8. Mellman, I., Plutner, H., Steinman, R. M., Unkeless, J. C. & Cohn, Z. A. *J. Cell Biol.* **96**, 887–895 (1983).
9. Young, J. D.-E., Unkeless, J. C., Young, T. M., Mauro, A. & Cohn, Z. A. *Nature* **306**, 186–189 (1983).
10. Young, J. D.-E., Unkeless, J. C., Kaback, H. R. & Cohn, Z. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1636–1640 (1983).
11. Green, S. A., Plutner, H. & Mellman, I. *J. Biol. Chem.* **260**, 9867–9874 (1985).
12. Holmes, K. L., Palfree, R. G. E., Hämmerling, U. & Morse, H. C., III *Proc. natn. Acad. Sci. U.S.A.* **82**, 7706–7710 (1985).
13. Hibbs, M. L., Hogarth, P. M. & McKenzie, F. C. *Immunogenetics* **22**, 335–348 (1985).
14. Kozak, C. A., Davidson, W. F. & Morse, H. C., III *Immunogenetics* **19**, 163–168 (1984).
15. Jones, B., Tite, J. P. & Janeway, C. A. Jr *J. Immun.* **136**, 348–356 (1986).
16. Young, R. A. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1194–1198 (1983).
17. von Heijne, G. *J. molec. Biol.* **173**, 243–251 (1984).
18. Kozak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).
19. Mostov, K. E., Freidlander, M. & Blobel, G. *Nature* **308**, 37–43 (1984).
20. Williams, A. F., Barclay, A. N., Clark, M. J. & Gagnon, J. in *Gene expression during normal and malignant differentiation* (eds Andersson, L. C., Gahrberg, C. G. & Ekblom, P.) (Academic, 1985).
21. Hemperly, J. J., Murray, B. A., Edelman, G. M. & Cunningham, B. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3037–3041 (1986).
22. D'Eustachio, P., Pratcheva, D., Marcu, K. & Ruddle, F. H. *J. exp. Med.* **151**, 1545–1550 (1980).
23. MacInnes, J. I., Morris, V. L., Flintoff, W. F. & Kozak, C. A. *Virology* **132**, 12–25 (1984).
24. Festenstein, H. *Transplant. Proceed.* **8**, 339–342 (1976).
25. Ullrich, A. *et al. Nature* **309**, 418–425 (1984).
26. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
27. Dale, R. M. K., McClure, B. A. & Houchins, J. P. *Plasmid* **13**, 31–40 (1985).
28. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
29. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

Loss of mouse fibroblast cell response to phorbol esters restored by microinjected protein kinase C

Gabriella Pasti*, Juan-Carlos Lacal†, Barbour S. Warren*, Stuart A. Aaronson† & Peter M. Blumberg*

* Molecular Mechanisms of Tumor Promotion Section, Laboratory of Cellular Carcinogenesis and Tumor Promotion, and † Laboratory of Cellular and Molecular Biology, National Cancer Institute, Bethesda, Maryland 20892, USA

The phorbol esters in addition to being among the most potent mouse skin tumour promoters profoundly affect many different biological systems^{1,2}. It is postulated that they act through activation of protein kinase C (ref. 3–5), but substantial heterogeneity in their pharmacological and binding behaviour in some systems⁶ has caused concern about whether this is their only target. Evidence linking protein kinase C activation with biological responses to the phorbol esters includes similarity in structure–activity relations for binding and response⁷; *in vitro* phosphorylation of specific proteins by protein kinase C at the same sites at which phorbol ester treatment induces phosphorylation in intact cells³; and corre-

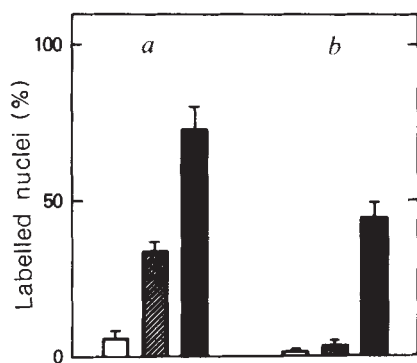


Fig. 1 Blocking of the mitogenic response of Swiss 3T3 cells to phorbol esters by pretreatment with phorbol ester. Control cells (a) or cells previously incubated with phorbol esters (b) were treated with no addition (□), 200 nM PDBu (▨) or 10% calf serum (■), and DNA synthesis measured by autoradiography. Data represent the mean $\pm$ s.e.m. for 5 experiments, with 200 cells counted for each experimental condition in each experiment.

Methods. Swiss 3T3 cells were plated at 1.2×10^5 cells per 35-mm tissue culture dish in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% calf serum (CS, Hyclone) and incubated at 37 °C in 95% air: 5% CO₂. Two days after plating, the cells were refed with fresh DMEM containing 10% CS. After an additional 2 days, the plates were washed 5 $\times$ with DMEM, and DMEM containing 1% CS was added back in the presence or absence of 200 nM PDBu. Three days later all the cultures were washed 5 times with 1 mg ml⁻¹ bovine serum albumin (BSA) in phosphate-buffered saline to remove residual PDBu. Conditioned DMEM containing 1% CS was added back together with 2 μ Ci ml⁻¹ ³H-thymidine (NEN) and 200 nM PDBu or 10% CS. After incubation for a further 24 h, the plates were washed and subjected to autoradiography as described^{15,16}, and labelled nuclei counted.

lation in certain cell types between down regulation of protein kinase C on chronic phorbol ester treatment and loss of cellular responsiveness to the phorbol ester⁸⁻¹⁰. Here we report that microinjection of purified protein kinase C into Swiss 3T3 fibroblasts pretreated with the phorbol ester phorbol 12,13-dibutyrate (PDBu) restores the mitogenic response of the cells to PDBu, directly establishing the involvement of protein kinase C in this response.

Quiescent cultures of Swiss 3T3 cells were obtained in the presence of low serum (Fig. 1 legend). Under these conditions addition of 200 nM PDBu stimulates DNA synthesis, measured by autoradiography, in $34 \pm 3\%$ (mean $\pm$ s.e.m.) of the cells compared with $83 \pm 7\%$ (mean $\pm$ s.e.m.) of cells stimulated with 10% serum (Fig. 1a). Pretreatment of the cultures for 3 days with 200 nM PDBu markedly reduces the number of cells responding to PDBu (Fig. 1b). The number of cells responding to 10% serum was also reduced although by a smaller percentage perhaps because the phorbol ester pretreatment resulted in higher cell density (protein per plate = 144% of controls).

Similar results were obtained when DNA synthesis was measured by total ³H-thymidine incorporation. In control cells, stimulation of ³H-thymidine incorporation by PDBu was $35 \pm 6\%$ (mean $\pm$ s.e.m.) of that by 10% serum (5 experiments). In PDBu pretreated cells PDBu caused no stimulation, whereas 10% serum stimulated ³H-thymidine incorporation to $78 \pm 12\%$ (mean $\pm$ s.e.m.) of the level for cells that were not pretreated.

To determine the effect of restoration of protein kinase C on the mitogenic response, cells were microinjected with $4-8 \times 10^5$ molecules per cell of active protein kinase C. This amount was comparable to the initial kinase levels in control cells of $3-6 \times 10^5$ (ref. 8). To control for effects caused by microinjection, results were compared with those for cells microinjected with carrier protein only. Microinjected protein kinase C mitogenically stimulates phorbol ester pretreated cells provided 200 nM PDBu was present (Fig. 2). No stimulation was seen in the absence of PDBu in either control or pretreated cells, confirming that the

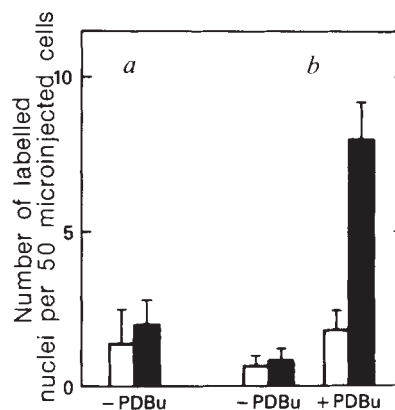


Fig. 2 Stimulation of mitogenesis by microinjected protein kinase C in phorbol-ester pretreated cells. Control cells (a) or cells pretreated with phorbol esters (b) were microinjected with protein kinase C (■) or with carrier solution (□). They were then incubated with ³H-thymidine for 24 h in the absence or presence of 200 nM PDBu, and labelled nuclei in the microinjected area were counted. Data represent the mean $\pm$ s.e.m. for 5 experiments using two separate preparations of protein kinase C. In each experiment, 50-100 cells were microinjected under each experimental condition.

Methods. Swiss 3T3 cells were grown in 60-mm tissue-culture dishes and treated as described (Fig. 1 legend). Protein kinase C was purified to homogeneity as described by Jeng *et al.*¹⁷. After the final purification step the enzyme was concentrated in the presence of fatty acid-free BSA (Sigma), and the buffer was changed to 130 mM KCl. The final BSA concentration was 15 mg ml⁻¹. The concentrations of protein kinase C in the two preparations used were 2.2 and 4.1 mg ml⁻¹, respectively. The specific activities were 1.6 and 1.2 μ mol ³²P transferred mg⁻¹ min⁻¹, respectively. Activity in the absence of Ca²⁺ and phospholipid was <1% of that in its presence. The enzyme was stored for no more than 3 days at 0 °C before use. During this period, activity decreased no more than 20%. Volumes of 5×10^{-14} l were usually microinjected¹⁶ using a Leitz micromanipulator coupled to a Zeiss IM35 microscope under a F-Achomat LD 32/0.4 PH1 lens (320 $\times$ total). Pipettes were made from borosilicate glass tubes of 0.7-1 mm outer diameter (Kimble) using a vertical pipette puller Model 720 (D. Kopf Instruments, Tujunga, California). Treatment of tips and loading of pipettes by suction from the tips was as described previously^{16,18}. Labelling of microinjected control cells incubated in the presence of PDBu was not determined because the results would have been obscured by the mitogenic response of uninjected cells within the microinjected area.

mitogenic effect is phorbol ester-dependent. The absolute level of mitogenic stimulation, 17% of nuclei labelled, was lower than that observed for stimulation of control cells by PDBu. This difference may be more apparent than real, however, because these results are normalized to the total number of microinjected cells. Swiss 3T3 cells are somewhat fragile to the microinjection procedure under our culture conditions and ~40% of the cells lysed after injection either with carrier protein or with enzyme. The number of labelled nuclei per surviving microinjected cell would therefore be larger.

These studies provide direct evidence for the role of protein kinase C in the mitogenic response of Swiss 3T3 cells to the phorbol esters. The role of protein kinase C in mitogenesis is of particular interest because there is evidence that some growth factors and oncogene products induce enhanced phosphatidylinositol turnover, activating protein kinase C (ref. 11), although apparently protein kinase C cannot account for all their mitogenic activity¹⁰. Our results do not support the suggestion of Collins and Rozengurt¹⁰, based on the kinetics of recovery of response in the presence of cycloheximide, that phorbol ester pretreatment imposes an additional postreceptor block to responsiveness, but this difference may reflect the difference in stimulation conditions.

A similar strategy may be applicable to other systems, for example, cell-cell communication¹², in which desensitization is observed in response to chronic phorbol ester treatment. Micro-injection has been used to demonstrate a role for protein kinase C in two other systems: enhancement of calcium current in neurones of *Aplysia*¹³, and K⁺ ion conductance in photoreceptors of *Hermisenda*¹⁴.

We thank Dr Stuart H. Yuspa for critical reading of the manuscript.

Received June 10; accepted 22 August 1986.

1. Slaga, T. J. (ed.) *Mechanisms of Tumor Promotion* (Raven, New York, 1984).
2. Blumberg, P. M. *Crit. Rev. Toxicol.* **8**, 153-234 (1981).
3. Nishizuka, Y. *Nature* **308**, 693-698 (1984).
4. Nield, J. E., Kuhn, L. J. & Vandenbark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 36-40 (1983).
5. Leach, K. L., James, M. & Blumberg, P. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4208-4212 (1983).
6. Blumberg, P. M. *Biochem. Pharmacol.* **33**, 933-940 (1984).
7. Blumberg, P. M. *et al. Mechanisms of Tumor Promotion* Vol. 3 (ed. Slaga, T. J.) 143-184 (CRC, Boca Raton, Florida, 1984).
8. Collins, M. K. L. & Rozengurt, E. *J. cell. Physiol.* **112**, 42-50 (1982).
9. Rodriguez-Pena, A. & Rozengurt, E. *Biochem. biophys. Res. Commun.* **120**, 1053-1059 (1984).
10. Collins, M. K. L. & Rozengurt, E. *J. cell. Physiol.* **118**, 133-142 (1984).
11. Berridge, M. J. *Biotechnology* 541-546 (1984).
12. Enomoto, T. & Yamaski, H. *Cancer Res.* **45**, 3706-3710 (1985).
13. DeRiemer, S. A., Strong, J. A., Albert, K. A., Greengard, P. & Kaczmarek, L. K. *Nature* **313**, 313-316 (1985).
14. Farley, J. & Auerbach, S. *Nature* **319**, 220-223 (1986).
15. Adams, R. L. P. *Expl Cell Res.* **56**, 49-54 (1969).
16. Lacal, J. C., Srivastava, S. K., Anderson, P. S. & Aaronson, S. A. *Cell* **44**, 609-617 (1986).
17. Jeng, A. Y., Sharkey, N. A. & Blumberg, P. M. *Cancer Res.* **46**, 1966-1971 (1986).
18. Graessmann, A., Graessmann, M. & Müller, C. *Meth. Enzym.* **65**, 816-825 (1980).

The *v-fms* oncogene induces factor independence and tumorigenicity in CSF-1 dependent macrophage cell line

Esther F. Wheeler, Carl W. Rettenmier, A. Thomas Look & Charles J. Sherr

Department of Tumor Cell Biology, St Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, Tennessee 38105 USA

The McDonough strain of feline sarcoma virus (SM-FeSV) transforms fibroblast cell lines in culture and produces fibrosarcomas in domestic cats^{1,2}. SM-FeSV does not induce haematopoietic malignancies in spite of the fact that its viral oncogene, *v-fms*, codes for a glycoprotein related to the receptor for the mononuclear phagocyte colony stimulating factor, CSF-1 (refs 3, 4). The *v-fms*-coded polypeptide includes the complete extracellular domain of the *c-fms* proto-oncogene product^{5,6} and retains the ability to bind CSF-1 specifically⁴. The two molecules have very similar sequences except at their extreme carboxyl terminal ends where 40 amino acids of the *c-fms*-coded glycoprotein are replaced by 11 unrelated residues in the *v-fms* product⁵. Autophosphorylation of the *c-fms* gene product on tyrosine is enhanced by CSF-1 addition³, whereas phosphorylation of the *v-fms*-coded glycoprotein appears to be constitutive⁴. We now show that introduction of the *v-fms* gene into simian virus-40 (SV40)-immortalized, CSF-1 dependent macrophages renders them independent of CSF-1 for growth and tumorigenic in nude mice. These factor-independent cell lines express unaltered levels of the *c-fms* product which is down-modulated in response to either CSF-1 or the tumour promoter 12-*O*-tetradecanoyl-phorbol-13-acetate (TPA). The induction of factor independence by a non-autocrine mechanism suggests that the *v-fms* product is an unregulated kinase that provides growth stimulatory signals in the absence of ligand.

The murine macrophage cell line BAC1 was derived from adherent BALB/c × A.CA F₁ mouse spleen cells by transfection with origin-defective SV40 DNA⁷. The 2F5 subclone⁸, chosen for its inability to grow in the absence of CSF-1, produces

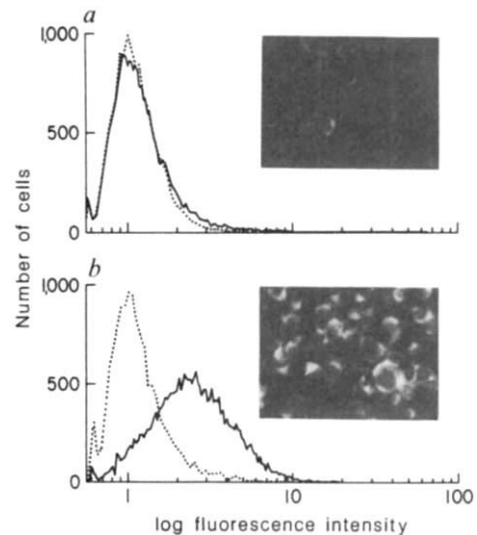


Fig. 1 Expression of *v-fms*-coded antigens in SM-FeSV-infected BAC1.2F5 macrophages. SM-FeSV-infected cells maintained for 5 passages in *a*, the presence or *b*, the absence of CSF-1 were tested for expression of *v-fms*-coded epitopes identified by the rat monoclonal antibody SM2.6.3^{11,13}. Solid line, the flow cytometric fluorescence profile recorded after binding of SM2.6.3 to the surface of viable cells; dotted line, the control fluorescence profile obtained with an irrelevant isotype-matched rat myeloma protein. The insets in each panel show photomicrographs (400 × magnification) of fixed, permeabilized cells incubated with the same monoclonal antibody and subsequently reacted with an affinity purified, fluorescein-labelled goat antibody to rat immunoglobulin G (IgG). Note the single positive cell in the field of factor dependent cells shown in *a*. The frequency of positive cells in these cultures was <2% of the total population (see Table 1). BAC1.2F5 cells studied in parallel showed background fluorescence, similar to that observed for the majority of cells in *a*.

Methods. Indirect immunofluorescence labelling of viable cells and flow cytometric analysis were performed as described¹¹, except that cells scraped from culture dishes were preincubated in medium containing 100 mg ml⁻¹ bovine IgG before staining, to prevent nonspecific binding of the monoclonal antibody to Fc receptors. Immunofluorescence staining of methanol-fixed cells was as described¹³, except that the slides were pretreated with buffer containing 100 mg ml⁻¹ of bovine IgG before reaction with the monoclonal antibody. The insets in *a* and *b* are matched exposures of SM-FeSV-infected FD and FI cells, respectively, photographed in the same experiment.

lysozyme, collagenase and esterase, secretes interleukin-1, expresses Fc receptors, and is capable of Fc-mediated phagocytosis. BAC1.2F5 cells were infected with high titre, helper-free SM-FeSV pseudotype virus⁶ prepared by transfection of cloned, replication-defective SM-FeSV DNA⁹ into the packaging NIH-3T3 cell line, Ψ2, which contains a defective Moloney murine leukaemia provirus that provides viral replicative functions but cannot package its own RNA¹⁰. No factor-independent colonies were obtained from mock-infected BAC1.2F5 cells. SM-FeSV-infected cells grown for seven days in the absence of CSF-1 gave ~0.1% factor-independent (FI) colonies. These were grown and maintained in the absence of CSF-1. Virus-infected factor-dependent (FD) cell lines were maintained in L-cell conditioned medium containing CSF-1 and passaged at 5-day intervals.

SM-FeSV-infected macrophages were plated in semisolid medium with or without CSF-1. Uninfected BAC1.2F5 cells formed colonies in agar only in the presence of CSF-1 (Table 1); FI cells formed colonies (6.9%) in the absence of CSF-1; as did FD cells at passage 3 and 5 after infection, but at low frequency (0.1-0.3%). The plating efficiency of FD cells was greatly increased (6.0%) after 14 passages. When factor-

Table 1 Efficiency of colony formation of SM-FeSV-infected BAC1.2F5 cells

	Colony forming efficiency (% of plated cells)	
	+CSF-1	-CSF-1
Uninfected BAC1.2F5 cells	7.0	<0.01
FD cells:		
passage 3	6.8	0.1
passage 5	4.6	0.3
passage 14	10.9	6.0
FI cells:		
passage 5	9.3	6.9
passage 10	13.4	4.2

A plasmid containing biologically active SM-FeSV proviral DNA⁹ was transfected into Ψ 2 cells¹⁰, and transformed Ψ 2 subclones were derived from single cells by endpoint dilution in microtitre culture dishes. Supernatants from a subclone producing 2×10^7 helper-free SM-FeSV focus-forming units (FFU) per ml on NIH-3T3 cells⁶ were used to infect 3×10^5 BAC1.2F5 cells at 10 FFU per cell. Cells were grown in Dulbecco's minimal medium containing 15% fetal calf serum and 25% murine L-cell conditioned medium containing CSF-1. Two days after infection the culture was divided and infected cells were continuously maintained in the presence or absence of CSF-1. In the absence of CSF-1 most of the infected cells failed to survive, but about 0.1% of the unpassaged cells gave rise to factor independent colonies which were grown up and maintained in the absence of CSF-1 (FI cells). Parallel cultures maintained in medium containing the growth factor (FD cells) gave rise to FI cells at a progressively increased frequency when shifted to factor-free medium. Cells from both infected populations were plated at various passage levels after infection in 60 mm culture dishes in 0.3% Noble agar containing complete medium supplemented with or lacking CSF-1. In each experiment duplicate sets of cultures were established containing 10^3 , 10^4 and 10^5 cells per dish. The efficiency of colony formation was independent of the cell density and represents an average of triplicate determinations. Continuously growing colonies (>100 cells) were scored three weeks after plating. Colonies picked from SM-FeSV-infected cells plated in the absence of CSF-1 gave rise to cell lines which were indistinguishable in their properties from the original FI cell line derived in liquid medium.

independent agar colonies from FD cells at passages 5 and 14 were picked and cultured in liquid medium without CSF-1, the resulting subclones were indistinguishable from unpassaged FI cells directly after infection. Thus SM-FeSV-infected cells maintained in medium containing CSF-1 progressively gave rise to FI cells. Because the SM-FeSV was produced from the Ψ 2 packaging cell line, this was not due to virus spread. No RNA-dependent DNA polymerase or focus-forming activity was found in the culture fluids, so no virus was produced by infected cultures.

From Southern blot analysis with a *v-fms* probe, early passage FD cells contained an average of 0.5–1.0 proviral DNA copies per cell, integrated polyclonally in host DNA, but late-passage FD cells, and FI subclones derived from single cells contained multiple (~2–4) proviruses per cell (data not shown). Most FD cells must have acquired proviral DNA after infection, but only those with multiple proviral insertions appeared to have a proliferative advantage, even in the presence of CSF-1.

Conditioned medium from FI cultures was tested using uninfected BAC1.2F5 cells to exclude the possibility that FI cells produced CSF-1. BAC1.2F5 cells die within 7 days in the absence of CSF-1. No CSF-1 activity was detected, so the proliferation of FI cells is not due to an autocrine mechanism. Northern blot analysis using a CSF-1 complementary DNA probe confirmed that BAC1.2F5, FD and FI cells did not express CSF-1 messenger RNA.

In fibroblasts infected with SM-FeSV the *v-fms*-coded glycoprotein, gp140^{*v-fms*}, is required at high levels on the cell surface for transformation^{11,12}. Only a low percentage of unpassaged SM-FeSV-infected FD cells gave rise to FI colonies when CSF-1

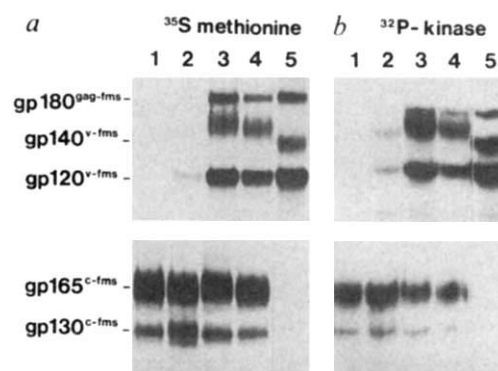


Fig. 2 Synthesis (a) and tyrosine kinase activities (b) of *v-fms* (top panels) and *c-fms* (bottom panels) gene products in infected mouse cells. The lanes are the same in each panel. Lane 1, uninfected BAC1.2F5 cells; lane 2, SM-FeSV-infected FD cells maintained in medium containing CSF-1 for 5 passages; lane 3, infected FD cells maintained for 14 passages; lane 4, SM-FeSV-infected FI cells at passage 5; lane 5, SM-FeSV-transformed mouse NIH-3T3 fibroblasts. The mobilities of the various forms of *v-fms* and *c-fms* gene products are noted at the left margins. Due to differences in oligosaccharide processing, the mature form of the *v-fms*-coded glycoprotein (gp140^{*v-fms*}) in the mouse macrophages has a greater apparent molecular mass than that expressed in NIH-3T3 cells. Matched exposures of the corresponding autoradiographs are shown in upper and lower panels.

Methods. To avoid down modulation of CSF-1 receptors¹⁸ in cells grown in the presence of CSF-1, all cell lines were shifted to factor-free medium 18 h before the experiment. a, Cells were metabolically radiolabelled with ³⁵S-methionine for 90 min and then lysed in RIPA buffer (50 mM Tris HCl, pH 7.4, containing 150 mM NaCl, 20 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate and 0.1% sodium dodecyl sulphate (SDS)) with 2% Aprotinin and 1 mM phenylmethylsulphonyl fluoride as protease inhibitors. Unlabelled cell lysates for protein kinase assays were prepared in parallel. The lysates were incubated first with antibodies that react with the *v-fms*-coded glycoproteins but do not bind the product of murine *c-fms* gene¹³. After recovery of the immune complexes using *Staphylococcus aureus* protein-A-Sepharose as immunoadsorbant, the cleared lysates were incubated with a rabbit antiserum to a recombinant *v-fms*-coded polypeptide expressed in bacteria that cross reacts with the murine CSF-1 receptor^{3,14}. b, Unlabelled immune complexes were incubated in the presence of [γ -³²P]ATP and divalent cation as described^{16,17}. Radiolabelled products in washed immunoprecipitates were separated by SDS gel electrophoresis and detected by fluorography (a) or autoradiography (b).

was removed from the medium, suggesting that most of these cells produced insufficient amounts of the glycoprotein to become factor independent. We analysed the infected cultures by fluorescence-activated flow cytometry using a monoclonal antibody to the extracellular, amino-terminal domain of the *v-fms*-coded polypeptide^{11,13}. FD cells at passage 5, like uninfected BAC1.2F5 cells, did not express the *v-fms*-coded glycoprotein (Fig. 1a). In contrast, FI cells at passage 5 expressed relatively high levels of the glycoprotein (Fig. 1b). The FI cells were uniformly fluorescence-positive when assayed for cytoplasmic fluorescence (Fig. 1b, inset), whereas less than 2% of FD cells were fluorescent (Fig. 1a, inset). Thus, only a small proportion of early passage SM-FeSV-infected BAC1.2F5 cells express sufficient *v-fms* gene product to become factor independent.

We also examined the expression of the feline *v-fms* and murine *c-fms* gene products in the various cell lines. Cultures were metabolically radiolabelled with ³⁵S-methionine, lysed with detergent and sequentially immunoprecipitated; the first antibody was specific for the *v-fms* gene product¹³ (Fig. 2a, top), and the second, a rabbit antiserum to a recombinant *v-fms*-coded polypeptide, cross-reacts with the murine CSF-1 receptor^{3,14} (Fig. 2a, bottom). Because *v-fms* sequences were inserted into the open reading frame of the SM-FeSV *gag* gene, the primary

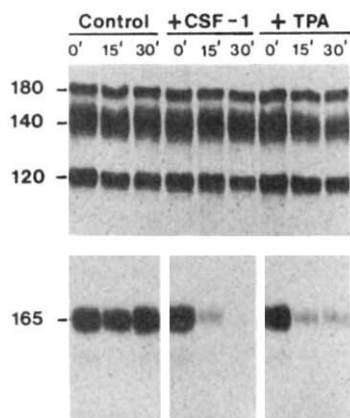


Fig. 3 Down modulation of the *c-fms* gene product in SM-FeSV-infected FI cells. Upper panel, *v-fms* gene products; lower panel, *c-fms* gene products. FI cells grown in factor-free medium were metabolically labelled for 15 min with ^{35}S -methionine and incubated for an additional 75 min in medium containing a 100-fold excess of unlabelled methionine. Following the 'chase' period, parallel tissue culture dishes were incubated in normal medium (control) or in medium containing either 3 nM CSF-1 or 1×10^{-7} M TPA for the indicated periods of time (0, 15, 30 min). Cells were lysed in RIPA buffer. The clarified lysates were incubated first with antibodies that react only with the *v-fms*-coded glycoproteins, and then with an antiserum to the *c-fms* product as in Fig. 2 legend. The immunoprecipitates were analysed by SDS gel electrophoresis. The sizes of the expected glycoproteins are given in kilodaltons (left). The figure gives data from quantitative sequential immunoprecipitations and shows matched exposures of the corresponding autoradiograms.

translation product is a glycosylated polypeptide of relative molecular mass (M_r) 180,000 (180K), (gp180^{gag-fms}) which is proteolytically cleaved, yielding an immature form of the *v-fms*-coded glycoprotein (gp120^{v-fms}), which undergoes processing and appears at the cell surface as a glycoprotein of greater M_r , gp140^{v-fms} (refs 13, 15). Uninfected BAC1.2F5 cells did not express *v-fms* gene products (Fig. 2a, top, lane 1). SM-FeSV-infected FD cells at passage 5 produced low levels of these glycoproteins (lane 2), but after serial subculturing in the presence of CSF-1, these cells (passage 14; lane 3) expressed comparable levels of the viral oncogene products to FI cells (lane 4) and SM-FeSV-transformed NIH-3T3 cells (lane 5).

Uninfected and infected BAC1.2F5 cells, whether CSF-1 dependent or not, expressed high levels of the murine *c-fms* gene product (Fig. 2a, bottom, lanes 1–4). The glycoprotein gp130^{c-fms} is an intracellular precursor of the mature 165K cell surface receptor, gp165^{c-fms} (ref. 14). SM-FeSV transformed NIH/3T3 cells did not express the *c-fms* product (lane 5). Immune complex protein kinase assays^{16,17} show that the *v-fms* (Fig. 2b, top) and *c-fms* (Fig. 2b, bottom) glycoproteins were phosphorylated in ratios proportional to the expression levels detected by metabolic radiolabelling. Phosphoamino acid analysis confirmed that *in vitro* phosphorylation was exclusively on tyrosyl residues. Subculturing of SM-FeSV-infected FD cells yields a population expressing high levels of the *v-fms*-coded glycoproteins without affecting synthesis of the *c-fms* product.

Since both the *v-fms* and *c-fms* products were coexpressed in FI cells, we studied down-modulation of the *c-fms* gene product¹⁸ after addition of CSF-1 or TPA. Cells labelled with ^{35}S -methionine were exposed to CSF-1 or TPA for brief periods, and the *v-fms* and *c-fms* products were immunoprecipitated from detergent lysates and analysed on gels. Turnover of gp140^{v-fms} was unaffected by CSF-1 or TPA, whereas turnover of gp165^{c-fms} was greatly increased by either compound (Fig. 3). Thus, both CSF-1 and TPA down-modulate gp165^{c-fms} but not the transforming glycoprotein.

Since the SM-FeSV-infected FI cells did not produce CSF-1 or show constitutive down-modulation of their CSF-1 receptors, their conversion to factor independence was directly mediated by the *v-fms*-coded receptor kinase rather than an autocrine mechanism. Thus, although the *v-fms* gene product includes the ligand-binding domain of the CSF-1 receptor, it can function as an unregulated enzyme without ligand. This has not been formally shown in fibroblasts, since fibroblast cell lines susceptible to transformation by SM-FeSV produce CSF-1 (ref. 4). The structural differences between the carboxyl termini of the *v-fms* and *c-fms* products could be responsible for the transforming activity of the oncogene product⁵. The carboxyl terminus of the *v-fms*-coded glycoprotein lacks a distal tyrosine residue found in the *c-fms* product, so it may be that phosphorylation of the *c-fms* product on this tyrosine decreases kinase activity, as observed¹⁹ for pp60^{c-src}.

Late passage FD and FI cells both formed tumours when inoculated subcutaneously or intraperitoneally into nude mice (5×10^6 cells per animal), whereas uninfected SV40-immortalized BAC1.2F5 cells did not. Tumours appeared within 5 weeks after injection, spread haematogenously, and formed histiocytic sarcomas metastatic to liver, spleen, pancreas, lung and spinal cord. CSF-1 independent cell lines from tumour tissues had the histological appearance of macrophages, stained cytochemically for butyrate esterase, contained multiple copies of the integrated SM-FeSV provirus and expressed high levels of *v-fms*-coded glycoprotein at the cell surface. The *v-fms* receptor kinase therefore acts with a nuclear oncogene product to form tumours *in vivo*. Critical alterations in the kinase domain of the *c-fms* proto-oncogene might similarly contribute to haematopoietic malignancy.

Oncogenes encoding other tyrosine kinases, including *v-abl*, *v-src*, *v-fps/fes* and *v-erb-B*, can transform a variety of different cells of the myeloid and lymphoid lineages^{20–27}. The normal physiological functions of the *v-abl*, *v-src*, and *v-fps/fes* gene products remain obscure, but the ability of *v-erb-B* (a truncated EGF receptor gene)^{28,29} to transform erythroid progenitors suggests that viral tyrosine kinases can generate abnormal signals for cell proliferation. This raises the possibility that the *v-fms* gene product may convert other haematopoietic cell lines to factor independence, including those that do not express CSF-1 receptors.

We thank Dr E. Richard Stanley for BAC1.2F5 cells, Drs Ernest Kawasaki and Peter Ralph of Cetus Corporation for a CSF-1 cDNA clone and Shawn Kramer and Jean Elmendorf for technical assistance. This work was supported by grant CA-38187 (to C.J.S.), by Cancer Center grant CA-21765 from the National Cancer Institute and by the American Lebanese Syrian Associated Charities of St Jude Children's Research Hospital. E.F.W. was supported by a National Research Service Award (CA-09346) and C.W.R. by a Biomedical Research Support grant (RR-05584-21) from the NIH.

Received 7 July; accepted 22 September 1986.

- McDonough, S. K., Larsen, S., Brodey, R. S., Stock, N. D. & Hardy, W. D. Jr *Cancer Res.* **31**, 953–956 (1971).
- Sarma, P. S., Sharar, A. L. & McDonough, S. *Proc. Soc. exp. Biol. Med.* **140**, 1365–1368 (1972).
- Sherr, C. J. *et al. Cell* **41**, 665–676 (1985).
- Sacca, R., Stanley, E. R., Sherr, C. J. & Rettenmier, C. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3331–3335 (1986).
- Coussens, L. *et al. Nature* **320**, 277–280 (1986).
- Wheeler, E. F. *et al. J. Virol.* **59**, 224–233 (1986).
- Schwarzbaum, S., Halpern, R. & Diamond, B. *J. Immun.* **132**, 1158–1162 (1984).
- Morgan, C. J. & Stanley, E. R. *Biochem. biophys. Res. Commun.* **119**, 35–41 (1984).
- Donner, L., Fedele, L. A., Garon, C. F., Anderson, S. J. & Sherr, C. J. *J. Virol.* **41**, 489–500 (1982).
- Mann, R., Mulligan, R. C. & Baltimore, D. *Cell* **33**, 153–159 (1983).
- Roussel, M. F., Rettenmier, C. W., Look, A. T. & Sherr, C. J. *Molec. cell. Biol.* **4**, 1999–2009 (1984).
- Nichols, E. J., Manger, R., Hakomori, S., Herscovics, A. & Rohrschneider, L. R. *Molec. cell. Biol.* **5**, 3467–3475 (1985).
- Anderson, S. J., Furth, M., Wolff, L., Ruscetti, S. K. & Sherr, C. J. *J. Virol.* **44**, 696–702 (1982).
- Furman, W. L. *et al. Virology* **152**, 432–445 (1986).
- Anderson, S. J., Gonda, M. A., Rettenmier, C. W. & Sherr, C. J. *J. Virol.* **51**, 730–741 (1984).

16. Barbacid, M. & Lauver, A. V. *J. Virol.* **40**, 812-821 (1981).
17. Rettenmier, C. W., Chen, J. H., Roussel, M. F. & Sherr, C. J. *Science* **228**, 320-322 (1985).
18. Guilbert, L. J. & Stanley, E. R. *J. biol. Chem.* **261**, 4024-4032 (1986).
19. Cooper, J. A., Gould, K. L., Cartwright, C. A. & Hunter, T. *Science* **231**, 1431-1434 (1986).
20. Rosenberg, N. & Baltimore, D. *J. exp. Med.* **143**, 1453-1463 (1976).
21. Pierce, J. H. & Aaronson, S. A. *J. Virol.* **46**, 993-1002 (1983).
22. Adkins, B., Leutz, A. & Graf, T. *Cell* **39**, 439-445 (1984).
23. Kahn, P., Adkins, B., Beug, H. & Graf, T. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7122-7126 (1984).
24. Oliff, A., Agranovsky, O., McKinney, M. D., Murty, V. V. S. & Baehchwitz, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3306-3310 (1985).
25. Cook, W. D., Metcalf, D., Nicola, N. A., Burgess, A. W. & Walker, F. *Cell* **41**, 677-683 (1985).
26. Pierce, J. H. *et al. Cell* **41**, 685-693 (1985).
27. Anderson, S. M., Klinken, S. P. & Hankins, W. D. *Molec. cell. Biol.* **5**, 3369-3375 (1985).
28. Downward, J. *et al. Nature* **307**, 521-527 (1984).
29. Ullrich, A. *et al. Nature* **309**, 418-425 (1984).

Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots

Hagop Youssoufian*, Haig H. Kazazian Jr*, Deborah G. Phillips*, Sophia Aronist†, George Tsiftist†, Valerie A. Brown‡ & Stylianos E. Antonarakis*

* Genetics Unit, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

† Coagulation Laboratory, 1st Department of Pediatrics, Athens University School of Medicine, Athens, Greece

‡ Collaborative Research, Inc., Lexington, Massachusetts 02173, USA

Haemophilia A is a common disorder of blood coagulation caused by a deficiency of factor VIII¹. It is inherited as an X-linked recessive trait, and one-third of all cases are thought to result from *de novo* mutations². The clinical severity of haemophilia A varies markedly among different families and a subset of the patients with severe disease develop antibodies against factor VIII, called inhibitors³. Because of this heterogeneity, it is likely that many different molecular lesions result in haemophilia A. Indeed, of the nine mutations described to date, all appear to be unique changes⁴⁻⁶. However in this study of 83 patients with haemophilia A we have identified two different point mutations, one in exon 18 and one in exon 22, that have recurred independently in unrelated families. Each mutation produces a nonsense codon by a change of CG to TG, and each occurred *de novo* on the X-chromosome donated by the maternal grandfather. These observations strongly support the view that CpG dinucleotides are mutation hotspots.

Recent reports have described the isolation and partial sequence of the factor VIII gene and its expression *in vitro* from recombinant DNA clones⁷⁻¹⁰. The gene is 186 kilobases (kb) long and is divided into 26 exons, encoding a protein of 2,351 amino acids. The availability of cloned factor VIII gene fragments has made possible the study of molecular defects in patients with haemophilia A. Because of the clinical heterogeneity of the disorder, the large size of the factor VIII gene, and the pool of new mutations, it has been suggested that the number of genetic lesions producing haemophilia A may be very large^{4,5}.

We have analysed the genomic DNA of 83 unselected patients with haemophilia A who were referred to us for carrier detection and prenatal diagnosis. Our screening strategy for factor VIII mutations involved restriction analysis with *TaqI*, *SstI* and *EcoRI*, and successive hybridizations with two complementary DNA probes that span exons 1-12 (probe A) and 14-26 (probe BC) of the factor VIII gene. We detected 6 deletions and 10 point mutations using this strategy (ref. 5 and manuscript in preparation); 9 of the 10 point mutations were detected with *TaqI*, which has the dinucleotide CpG in its recognition sequence.

In an earlier study from our laboratory⁵, a patient with severe

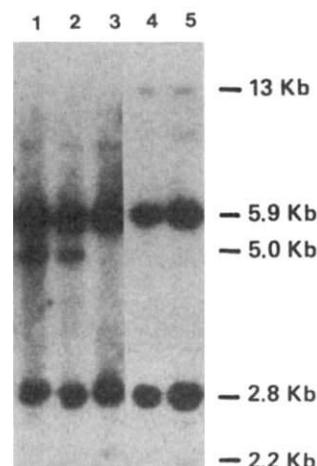


Fig. 1 Autoradiogram of DNAs from members of family K and family T digested with *TaqI* and hybridized to probe BC. Family K: lane 1, subject III-1; lane 2, subject II-2; lane 3, subject I-2; family T: lane 4, subject III-2; lane 5, subject II-2.

Methods. Genomic DNA was isolated from 10-15 ml of peripheral blood by standard techniques. DNA (5 µg) was digested to completion with *TaqI*. Gel electrophoresis, transfer to nitrocellulose filters, and hybridization with a ³²P-labelled cloned factor VIII DNA fragment (probe BC) were performed as previously described⁵. The following cloned factor VIII DNA fragments were used as probes: (1) probe A, a 1.7 kb *KpnI* cDNA fragment that spans exons 1-12; and (2) probe BC, a 6.5 kb *EcoRI* cDNA fragment that spans exons 14-26. These probes were kindly provided by Drs J. Toole and J. Wozney of Genetics Institute, Inc. The 2.2 kb fragments of subject II-2 (family K) and subject III-2 (family T) are difficult to see on this autoradiogram. Sizes of fragments are given in kb.

disease (factor VIII titre <1 U per 100 ml) and factor VIII inhibitors was shown to have a point mutation in exon 18 of the factor VIII gene, where a change of C to T in a CpG dinucleotide led to the generation of a nonsense codon. That family was of Anglo-Saxon origin, and the mutation was present in at least four generations. Upon screening the DNAs of our more recent patients, we observed a *TaqI* site alteration in an unrelated Greek family (family K) where the two affected brothers (subjects III-1 and III-2, Fig. 2a) had severe haemophilia A (factor VIII titre <1 U per 100 ml) and were inhibitor-negative. This change appeared identical to the mutation described previously.

Figure 1 shows a representative restriction analysis of genomic DNA from various members of this family digested with *TaqI*. Hybridization with probe BC revealed a new 5.0 kb fragment and absence of a 2.2 kb fragment in an affected male (subject III-1). The DNA of the mother, an obligate heterozygote (Fig. 2a, subject II-2), contained the 5.0 kb fragment as well as the 2.2 kb fragment. The patient's father, maternal grandfather and maternal grandmother (Fig. 2a; subjects II-1, I-1, and I-2, respectively) displayed the normal *TaqI* pattern, indicating that this mutation appeared *de novo* in one of the mother's X-chromosomes. No variant fragments were detected when the DNA was digested with *SstI*, *EcoRI*, or *BamHI*, or when probe A was used for hybridization (data not shown).

To determine the precise nature of the mutation responsible for the altered *TaqI* site, we used oligonucleotides to detect single base mismatches^{11,12}. A pair of oligonucleotides were made, each 18 nucleotides long, one having the normal sequence flanking the exon 18 *TaqI* site, the other differing from the normal by the substitution of a T for C in the *TaqI* site, the presumed mutant sequence. Digestion of genomic DNA with *HindIII* normally produces a 2.7 kb fragment which contains

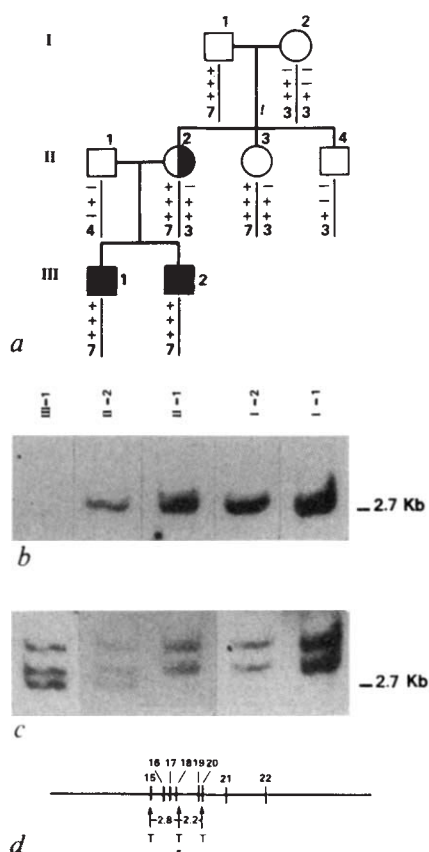


Fig. 2 Analysis of the factor VIII exon 18 mutation. *a*, Pedigree of family K (□, normal male; ■, haemophilic male; ○, normal female; ●, carrier female). Haplotypes of polymorphic restriction sites are indicated below each subject, in the order (top to bottom) *Bcl*I-factor VIII, *Bgl*II-factor VIII, *Bgl*II-DXS15, *Taq*I-DXS52. The presence or absence of a polymorphic restriction site is shown as + or -. In one case (DXS52) the alternative alleles are arbitrarily numbered. *b*, Oligonucleotide analysis of genomic DNA from members of family K digested with *Hind*III and hybridized to the normal exon 18 probe (GATTTCGATGGTATCTGCT). *c*, Hybridization of the same gel with an oligonucleotide corresponding to the mutant exon 18 probe (GATTTCGATGGTATCTGCT). *d*, Partial restriction map of factor VIII DNA cleaved with *Taq*I (T). Asterisk, location of the absent *Taq*I site in exon 18.

Methods. DXS15 cloned DNA was kindly provided by Dr K. Davies and DXS52 cloned DNA by Dr J. L. Mandel. Oligonucleotides (18 bases long) were chemically synthesized by Dr C. Riley of the Johns Hopkins University School of Medicine, end-labelled using T_4 -polynucleotide kinase and 32 P-ATP, further purified by electrophoresis on a 20% polyacrylamide sequencing gel, and the labelled probe extracted by the crush and soak method^{11,12}. Hybridization to dried agarose gel was performed at calculated melting temperature (T_m minus 4°C for 3 h, washed in 2× SSPE/0.1% SDS, then washed further in 2× SSPE at 2°C below T_m for 2 min), and autoradiographed. The agarose gel was re-used after stripping off the first probe by denaturation in 0.5 M NaOH/0.15 M NaCl for 30 min, and neutralization.

exon 18 of the factor VIII gene. The normal probe did not hybridize to the 2.7 kb fragment in *Hind*III-cleaved DNA from the affected male (Fig. 2*b*, subject III-1), although it did hybridize to the DNAs of unaffected family members. By contrast, the mutant probe hybridized to the 2.7 kb factor VIII fragment from the DNA of the affected male (subject III-1) and his mother (subject II-2), but not to that of the normal members of the family (Fig. 2*c*). Constant hybridizing sequences that were not factor VIII were seen in all individuals. We have thus shown that the basis of the exon 18 *Taq*I site mutation is a C to T change producing a nonsense codon, TGA, at position 1,960 in

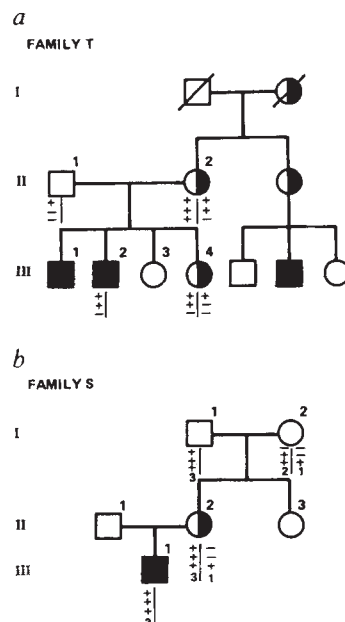


Fig. 3 Pedigrees of *a*, family T; *b*, family S. Both families show the factor VIII exon 22 mutation. Symbols as for Fig. 2*a*.

place of the normal CGA, which codes for arginine. The factor VIII protein produced by this mutant gene is presumably a truncated, inactive molecule.

To determine the parental origin of the new mutation in the patient's mother, we performed DNA polymorphism analysis on various family members. The DNA polymorphisms used were: *Bcl*I and *Bgl*II sites in the factor VIII gene; a *Bgl*II site adjacent to DXS15; and *Taq*I polymorphic alleles detected by probe DXS52^{5,13-15}. Analysis of these sites suggested that the X-chromosome of the patient was inherited from the maternal grandfather (Fig. 2*a*), and, therefore, the mutation to an abnormal factor VIII gene occurred in a germ cell of the maternal grandfather. The grandfather was 32 years old at the time his carrier daughter was conceived.

A different *Taq*I site alteration was found in both of two other unrelated families. The haemophilic subjects from each family have severe disease (factor VIII titres <1 U per 100 ml) but are inhibitor-negative. *Taq*I digestion of genomic DNA from subject III-2 of family T (Fig. 1) and subject III-1 of family S (data not shown) and hybridization to probe BC showed an abnormal 13 kb fragment and the absence of the normal 5.9 kb fragment that includes exon 22 sequences⁴. Female carriers of the mutation from family T (subjects III-4 and II-2, Fig. 3*a*) and subject II-2 from family S (Fig. 3*b*) also contained the 13 kb fragment in their DNA, but normal males from either family did not. Neither the maternal grandfather nor the maternal grandmother (subjects I-1 and I-2, respectively) of family S contained the abnormal fragment in their DNA, indicating that the mutation arose *de novo* within two generations.

We determined the precise nucleotide change by making a pair of oligonucleotides, each 18 nucleotides long, flanking the exon 22 *Taq*I site which differed in a single substitution of a T for C in the *Taq*I site. *Hind*III digestion of genomic DNA normally produces a 13 kb fragment which contains exon 22 sequences. The normal oligonucleotide probe failed to hybridize to the 13 kb fragment from the affected males (Fig. 4*a*) of family T (III-2) or family S (III-1), but the oligonucleotide bearing the mutant sequence showed specific hybridization to DNA of each affected individual (Fig. 4*b*) as well as to female carriers from each family (subjects II-2 and III-3 of family T and subject II-2 of family S). Therefore, the exon 22 mutation in both these

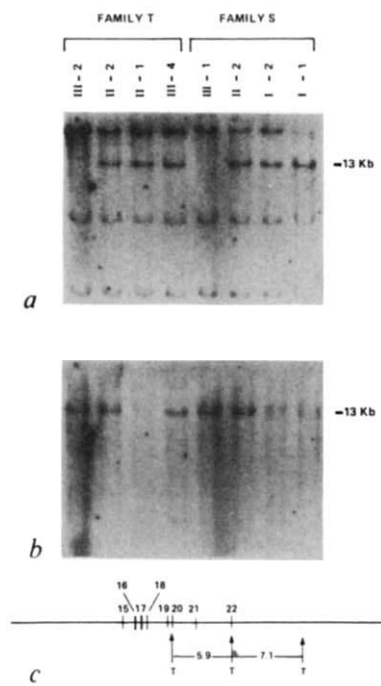


Fig. 4 Analysis of the factor VIII exon 22 mutations by oligonucleotide hybridization to *Hind*III-cleaved DNA from members of family T and family S. *a*, Hybridization with an 18-base long oligonucleotide corresponding to the normal exon 22 sequence (CTTATCGAGGAAATCCA). *b*, Hybridization of the same gel with an oligonucleotide corresponding to the mutant exon 22 sequence (CTTATTGAGGAAATCCA). *c*, Partial restriction map of factor VIII DNA cleaved with *Taq*I (T); asterisk, location of the absent *Taq*I site in exon 22. See Fig. 2 (legend) for methods.

families is due to a CGA to TGA change at codon 2,135 of the factor VIII gene, generating a nonsense codon.

By analysis of polymorphic restriction sites in the factor VIII gene and flanking it, we showed that the *de novo* mutation in family S arose on the X-chromosome of the maternal grandfather (subject I-1, Fig. 3b) who was 34 years old at the time his carrier daughter was conceived. It is interesting that both *de novo* mutations were associated with advanced paternal age; further data are needed to determine the significance of this observation.

In all four families, the haplotype of the two intragenic polymorphic sites (*Bcl*I and *Bgl*I) happened to be the same for chromosomes bearing the mutant gene. To confirm that these mutations were independent in origin and to exclude the possibility that the stated parentage was not correct in either family K or family S, we examined multiple, highly polymorphic loci that lie outside the factor VIII gene^{15,16}. We examined four such loci; in no case did the mothers in family K (Fig. 2a, subject II-2) or in family S (Fig. 3b, subject II-2) show a DNA fragment which was not consistent with stated parentage. The probability

of the stated parentage being correct is estimated for each patient's mother as 1×10^5 to 1.

These cases are two instances of recurrent mutation in man. The previously reported presence of a particular mutation in two very different ethnic groups and in two or more different β -globin gene frameworks provides strong circumstantial evidence for recurrent mutation in various β -globin gene alleles, including β^S , β^E , and certain β -thalassaemia alleles¹⁷. Chehab *et al.* have also presented strong evidence for recurrence of a common Mediterranean β -thalassaemia allele¹⁸. However, unlike β -thalassaemia, there is no positive selection for mutations that produce haemophilia A and the life of each mutation is estimated to be two to four generations in a stationary population². Thus, without positive selection for such mutations, each new mutation would be expected to last 100–150 years. This means that every unrelated affected family should have a different mutant allele (barring recurrent mutation) and that study of these alleles should give an unbiased observation of mutation hotspots.

There is strong circumstantial evidence that CpG dinucleotides are hotspots for mutation in man. The proportion of CpG in sequenced genes is 37% of that expected¹⁹. Several authors have postulated that methylated cytosines could be mutation hotspots because 5-methylcytosine can spontaneously deaminate to thymine resulting in a C-T transition in DNA^{20–22}. Barker *et al.*²³ found an unusually high frequency of polymorphism at a number of loci in human DNA using restriction enzymes whose recognition sequences contain a CpG dinucleotide. Significantly, of the four point mutations producing haemophilia A reported before this study, three were CG-TG changes and one a CG-CA mutation^{4–6}. The latter would result from a C-T transition at the CpG on the anti-sense strand of DNA. Our observation of recurrent CG-TG mutations strongly supports the view that these dinucleotides are mutation hotspots. One can estimate the number of recurrences of the codon 1,960 or codon 2,135 mutations in human history given that (1) their recurrence frequency is comparable to that observed in our studies (about 1 in every 50 affected males); (2) the incidence of haemophilia A is 1 in 10,000 males in all racial groups; and (3) each mutation persists for an average of 100–150 years in the population. Using these assumptions, we estimate that over 1,000 individuals are alive today with different recurrent origins of the C-T mutation at codon 1,960 or of the C-T mutation at codon 2,135 of the factor VIII gene, and that the total number of recurrences of each type of mutation in man is perhaps 10,000. Even if this is an overestimate by a factor of 10, it is in striking contrast to recurrence frequencies of 5–10 predicted for the β -globin alleles mentioned earlier, β^S , β^E , and certain β -thalassaemia mutations¹⁷.

We thank Drs J. Toole and J. Wozney for the factor VIII clones, K. Davies for the DXS15 clone, J. L. Mandel for clone DXS52, Drs A. Chakravarti and H. Donis-Keller for helpful discussions, Ms C. E. Dowling for technical assistance and Ms E. Pasterfield for assistance in the preparation of this manuscript. This study was supported by NIH grants to S.E.A. and H.H.K. and a Johns Hopkins Institutional Grant to S.E.A.

Received 6 June; accepted 12 September 1986.

- Hougie, C. in *Haematology* 3rd edn (eds Williams, W. J., Beutler, E., Erslev, A. J. & Lichtman, M. A.) 1381–1399 (McGraw-Hill, New York, 1983).
- Haldane, J. B. S. *J. Genet.* **31**, 317–326 (1935).
- Gill, F. M. in *Factor VIII Inhibitors* (ed. Hoyer, L. W.) 19–29 (Liss, New York, 1984).
- Gitschier, J. *et al. Nature* **315**, 427–430 (1985).
- Antonarakis, S. E. *et al. New Engl. J. Med.* **313**, 842–848 (1985).
- Gitschier, J. *et al. Science* **232**, 1415–1416 (1986).
- Gitschier, J. *et al. Nature* **312**, 326–330 (1984).
- Wood, W. I. *et al. Nature* **312**, 330–337 (1984).
- Vehar, G. A. *et al. Nature* **312**, 337–342 (1984).
- Toole, J. J. *et al. Nature* **312**, 342–347 (1984).

- Wallace, R. B. *et al. Nucleic Acids Res.* **9**, 3647–3656 (1981).
- Studnicki, A. B. & Wallace, R. B. *DNA* **3**, 7–15 (1984).
- Antonarakis, S. E. *et al. Lancet* **i**, 1407–1409 (1985).
- Goodfellow, P. N., Davies, K. E. & Ropers, H. H. *Cytogenet. Cell Genet.* **40**, 296–352 (1985).
- Schumm, J. *et al. Cytogenet. Cell Genet.* **40**, 739 (abstract) (1985).
- Kazazian, H. H. Jr *et al. Am. J. hum. Genet.* **38**, 860 (1986).
- Orkin, S. H. & Kazazian, H. H. Jr *Am. J. Rev. Genet.* **18**, 131–171 (1984).
- Chehab, F. F., Honig, G. R. & Kan, Y. W. *Lancet* **i**, 3–5 (1986).
- Nussinov, R. *J. molec. Biol.* **149**, 125–131 (1981).
- Salser, W. *Cold Spring Harb. Symp. quant. Biol.* **42**, 985–1002 (1977).
- Bird, A. P. *Nucleic Acids Res.* **8**, 1499–1504 (1980).
- Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. *Nature* **274**, 775–780 (1978).
- Barker, D., Schafer, M. & White, R. *Cell* **36**, 131–138 (1984).

The structure of β -lactoglobulin and its similarity to plasma retinol-binding protein

M. Z. Papiz*†, L. Sawyer*‡§, E. E. Eliopoulos||, A. C. T. North||, J. B. C. Findlay¶, R. Sivaprasadarao¶, T. A. Jones*, M. E. Newcomer* & P. J. Kraulis*

* Department of Applied Chemical Sciences, Napier College, Edinburgh EH10 5DT, UK

‡ Department of Biochemistry, University of Edinburgh, Edinburgh EH8 9XD, UK

|| Astbury Department of Biophysics, University of Leeds, Leeds LS2 9JT, UK

¶ Department of Biochemistry, University of Leeds, Leeds LS2 9JT, UK

* Department of Molecular Biology, Box 590, BMC University of Uppsala, Uppsala, Sweden

Since its first isolation¹, bovine β -lactoglobulin (BLG) has been an enigma: although it is abundant in the whey fraction of milk, its function is still not clear. The results of the many physico-chemical studies on the protein need a structural interpretation. We report here the structure of the orthorhombic crystal form of cow BLG at pH 7.6, at a resolution of 2.8 Å. It has an unusual protein fold, composed of two slabs of antiparallel β -sheet, which shows a remarkable similarity to plasma retinol-binding protein. A possible binding site for retinol in BLG has been identified by model-building. This suggests a role for BLG in vitamin A transport and we have discovered specific receptors for the BLG-retinol complex in the intestine of neonate calves.

BLG is found in the milk of many animals including bovine species, sheep, deer, horse, dog, pig, dolphin² and possibly man. Usually the protein exists as a dimer and many species appear to possess genetic variants. In ruminants, the relative molecular weight of the subunit is ~18,000 corresponding to a polypeptide of ~162 residues. There are two disulphide bridges and, in the cow protein, a free cysteine in the sequence Cys-Gln-Cys- (119–121)³, variously reported as 119⁴, 121^{5,6} or an equal mixture of the two⁷. The properties of the protein have been reviewed^{4,8–10}. It is remarkably acid-stable, resisting denaturation at pH 2¹¹ and remaining intact until after passing through the stomach¹². As the pH is raised, it undergoes several conformational changes observed by optical rotatory dispersion, perhaps the most interesting occurring at about pH 7.5¹³. BLG is known to bind a variety of hydrophobic molecules^{14–16}.

Of the abundant crystal forms of BLG^{17,18}, four salted-out forms have been studied by X-ray crystallography at low resolution¹⁹. A preliminary report has been given²⁰ of higher resolution work on one, lattice Y, crystallised at pH 7.6. We now discuss the similarity between BLG and plasma retinol-binding protein (RBP), which has suggested a function for BLG.

RBP is the protein that transports retinol from its storage in the liver. It is secreted into the plasma where it binds to trans-thyretin (formerly prealbumin). The transfer of retinol from plasma to retina is mediated by a specific surface receptor^{21,22}. Human RBP contains 182 amino acids; the 2 Å resolution structure of the RBP-retinol complex has recently been solved²³ and refined²⁴.

We have determined the structure of BLG by standard crystallographic techniques (Fig. 1). We are confident that the strands are accurately positioned in our map, although one external loop and the N- and C-terminal residues remain poorly defined. Higher resolution data now being processed should reveal the remainder of the molecule. The molecule consists of

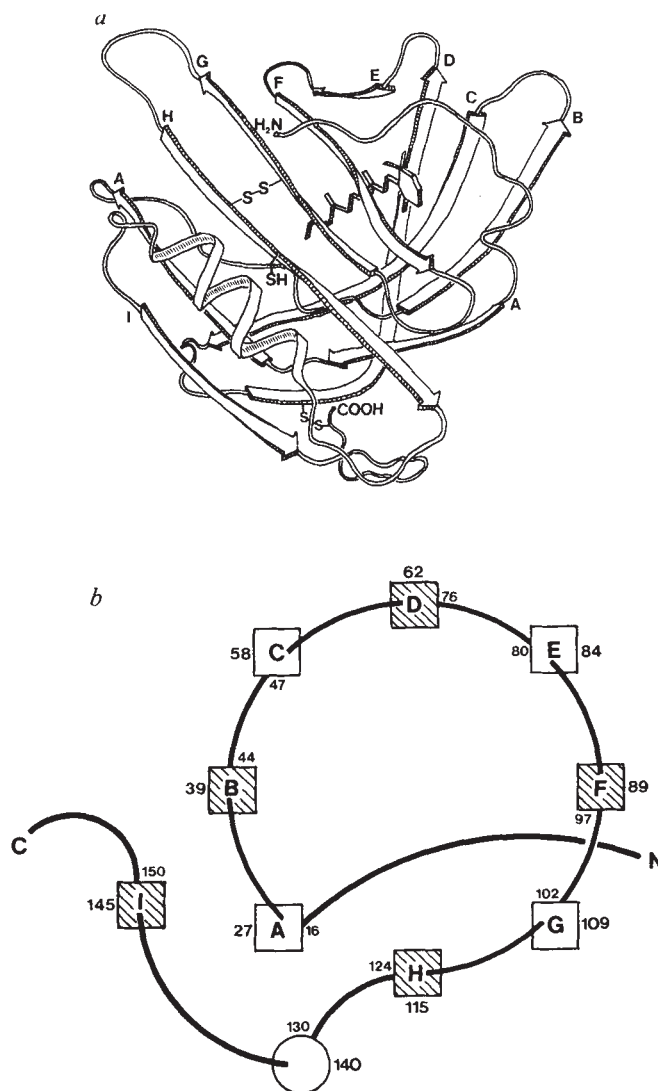


Fig. 1 a, Drawing of the β -lactoglobulin fold showing the position at which we propose that retinol binds. The arrows indicate the strands of β -sheet. The axis of the calyx runs approximately parallel to the extended isoprene chain with the far tip lying at the entrance from the solvent. b, The topology of the protein with the N-termini of the strands hatched to emphasize their antiparallel nature. A preliminary account of the preparation and crystallography of BLG is given in ref. 20.

anti-parallel β -sheet, formed by 9 strands wrapped round to form a flattened cone or calyx. The core of the molecule, an eight-stranded, antiparallel β -barrel, is an unusual structural motif which has been observed previously only in RBP²³, although catalase²⁵ has a barrel of similar topology, one connecting loop of which forms a separate domain. The strands A-I comprise residues 16–27, 39–44, 47–58, 62–76, 80–84, 89–97, 102–109, 115–124 and 145–150. Strands E, F, G, H, A and I are flanked by a 3-turn helix (130–140). The strands of the two sheets that form opposite sides of the calyx are orthogonal (ref. 26 and Fig. 2). Reverse turns occur at residues 44–47, 59–62, 78–81 and 84–88 but the turns at 98–102 and 111–115 are not true β -bends. Strand I is involved in the formation of the dimer by making anti-parallel interactions with the dyad-related strand. The interface between monomers also involves hydrophobic interactions between Ile 29 and Ile 147, and stacking of the imidazoles of symmetry-related His 146 residues. The free sulphhydryl, at position 121, is buried at the sheet-helix interface. The disulphide bond at 106–119 is clearly resolved and shows

† Present address: SERC Daresbury Laboratory, Warrington WA4 4AD, UK.

§ To whom correspondence should be addressed.

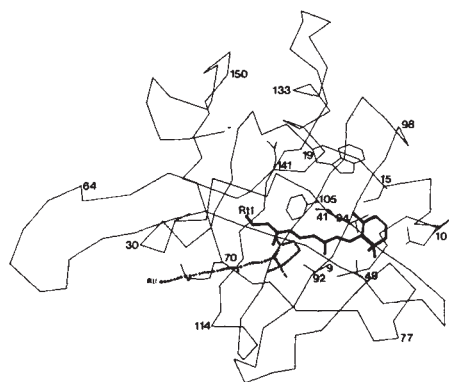


Fig. 3 Stereoscopic diagram of retinol (Rtl) model-built into β -lactoglobulin as described in the text. The side-chains of the BLG residues referred to are also shown as is the position of the retinol molecule (dotted) in RBP.

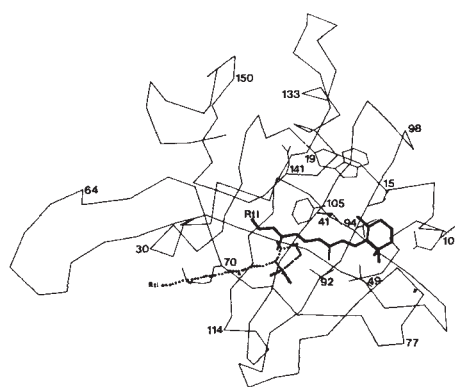
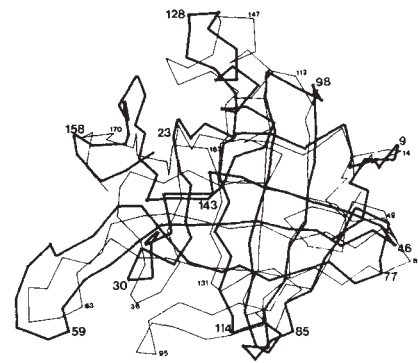
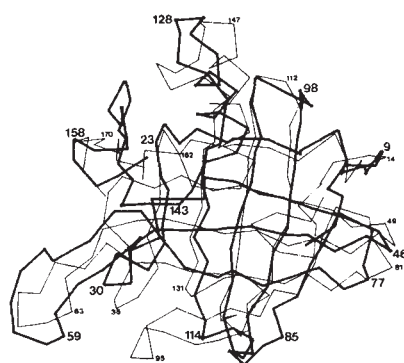


Fig. 2 Stereoscopic diagram of the superposition of the main chains of BLG (heavy) and RBP (light) viewed in such a way as to show the striking perpendicular packing of the β -sheets.



no evidence of equimolar amounts of cystine 106–119 and 106–121⁷. As the distance between the S_γ atoms of Cys 119 and Cys 121 is 10.5 Å, any interchange would be impossible without significant conformational flexibility. The other disulphide, 66–160, is not clear although it appears to be on the outer surface linking strand D to the C-terminal region.

BLG shows limited sequence homology to RBP, and human α_1 - and rat α_2 -microglobulins (refs 2, 27 and L. Rask and P. A. Peterson, personal communication). The sequence alignment of BLG and RBP, obtained with the program GAP²⁸ using a gap weight of 1 and a gap length of 0.1, shows 42 identities when 9 insertions or deletions are allowed. Only two pairs of regions have 4 consecutive identical residues (17–20 in BLG and 22–25 in RBP; 96–99 in BLG and 108–111 in RBP), and a third pair (79–83 in BLG and 83–87 in RBP) has 4 out of 5. However, the spatial similarity of the two structures is striking (Fig. 2) and when the backbone C_α coordinates were superimposed by the program HOMO^{29,30}, 129 out of the 162 residues were closely aligned, with an r.m.s. error of 2.76 Å. The greater length of the RBP chain is accommodated in larger surface loops, while the regular cores of the molecules are similar and in each case form a pocket open to the surface in which retinol binds in RBP.

Futterman and Heller¹⁵ studied the binding of retinol to BLG and RBP by fluorescence spectroscopy and suggested that the retinol binding site in RBP was non-polar. In RBP the fluorescence remained constant with time whereas BLG showed a slow decline. The retinol was fully protected against oxidation by liver alcohol dehydrogenase in the complex with RBP but not in that with BLG. It has been suggested³¹ that the BLG binding site makes the retinol more rigid than in free solution by encompassing the β -ionone ring and one or more adjacent carbon-carbon double bonds and that a tryptophan residue helps to form the binding site for the β -ionone moiety. Binding studies of retinylidene amine³² suggest that both BLG and RBP protect

the retinyl Schiff's base from hydrolysis and that in BLG there is a positively charged group near the carbon-nitrogen double bond. Interestingly, the dissociation constant, K_d , for the retinol-BLG complex³¹ is 2×10^{-8} M whereas that for the RBP complex³³ is only 1.7×10^{-7} M.

These studies are in agreement with our structural models. In RBP²⁴, the retinol molecule is surrounded by the protein except for the very tip of the hydroxyl group. We modelled the retinol-BLG complex, first transforming the coordinates of retinol from RBP to the BLG coordinate system. Interactive graphics^{34,35} gave a model in which the isoprene tail of retinol is fairly accessible to solvent because the loop at positions 83–85 in BLG is shorter than the corresponding one (at positions 91–98) in RBP. No tryptophan residue is close to the retinol although Trp 19, at the bottom of the calyx, lies at a distance of about 10 Å. Lys 70 may approach the retinol hydroxyl group. However, in BLG there appears to be an empty space further into the calyx and inaccessible to solvent, owing to changes from the residues that block this end of the cavity, in RBP. In particular, of 5 phenylalanine residues in RBP (at positions 15, 20, 45, 77 & 86) there are 3 (at positions 15, 20 and 45) that are replaced by smaller ones in BLG (Leu 10, Val 15 & Val 41); in addition, RBP Met 53 is changed to Thr 49 in BLG, RBP His 104 to BLG Val 92, RBP Ile 106 to BLG Val 94 and RBP Gln 117 to the larger Phe 105 in BLG. These changes allow the β -ionone ring of retinol to be fitted more deeply after minor conformational adjustments to some protein sidechains. The new binding site is partly formed by the sidechain of Trp 19. Most of the retinol is as well protected from the solvent in this conformation of BLG as in RBP, although the tip of the isoprene tail remains more exposed. It seems reasonable to imagine that some exchange is possible with the hydrophobic substrate binding pocket of ADH³⁶, which in the presence of NAD^+ could lead to oxidation. The retinol hydroxyl group is still within reach of the N_ϵ of Lys 70. Energy minimization of the retinol in this

deeper position reveal a better fit when the conjugated π -system is fully planar than when rotation is allowed about the 6-7 bond to permit closer approach to Trp 19. The fine structure in the UV spectrum of retinol when bound to BLG has been interpreted³⁷ as indicating a fully planar π -system. Thus, this model (Fig. 3) accounts for all the available chemical evidence, although details of the interactions between BLG and retinol will have to await the high resolution refined structure.

The fact that retinol binds to BLG with greater affinity than to RBP and our demonstration that it can be neatly accommodated within the BLG structure, prompted the speculation²⁰ that the biological function of the milk protein is involved with vitamin A transport. We monitored [³H]-retinol bound to BLG using both gel filtration and a transthyretin-affinity resin, but found no detectable interaction between BLG and transthyretin. However, [¹²⁵I]-labelled BLG-retinol complex binds specifically to purified microvilli prepared from one-week-old calf intestine. Total binding of [¹²⁵I]-labelled-BLG (5 nM) was reduced by 70% in the presence of 5 μ M unlabelled BLG. Specific binding was observed only in the lower segments of the small intestine and was not present in six-month old animals. The binding is dependent on both pH and ionic strength, and shows optima in the region of 6.8 and 0.1 M respectively. These observations are still preliminary, but they suggest that specific receptors for BLG which may be important in vitamin uptake are present in the small intestine of the neonate calf. The mechanism of uptake remains to be established. Neither RBP nor BLG inhibits the binding of the other to its specific receptor.

We thank Professor Sir David Phillips for drawing our attention to the similarity between β -lactoglobulin and RBP, Dr C. Holt for help with protein purification, Drs H. Muirhead and H. C. Watson, Bristol University, for use of X-ray photographic facilities, the Physics Department, Edinburgh University for use of their diffractometer and computing facilities and the SERC for financial support. E.E.E. thanks the University of Leeds for a studentship.

DNA conformation is determined by economics in the hydration of phosphate groups

Wolfram Saenger*, William N. Hunter† & Olga Kennard†

* Institut für Kristallographie Freie Universität Berlin, Takustr. 6, D-1000 Berlin 33, FRG

† University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK

Mixed sequence DNA can exist in two right-handed and one left-handed double helical conformations—A, B and Z¹⁻³. Under conditions of high water activity the B conformation prevails. If the water activity is reduced on addition of salt or organic solvents, transformation occurs to A-DNA or, in DNAs with alternating purine-pyrimidine sequences, to the left-handed Z-DNA. In crystal structure analyses of oligonucleotides, the free oxygen atoms of adjacent phosphate groups along the polynucleotide chain in B-DNA are found at least 6.6 Å apart and individually hydrated⁴ whereas they are as close as 5.3 Å in A-DNA and 4.4 Å in Z-DNA, and bridged by water molecules⁵⁻⁷. We suggest that this more economical hydration in A- and Z-DNA compared with B-DNA is the underlying cause of B $\rightarrow$ A and B $\rightarrow$ Z transitions.

To learn about possible driving forces for conformational transitions of DNA, we have investigated in detail the hydration of crystalline DNA fragments in the A, B and Z forms. The bases and sugars are equally well hydrated in all three forms. Ordered patterns of hydration have been observed in several oligonucleotide crystal structures, including the spines of hydration in the minor groove of the Z-hexamers with alternating G-C sequence⁶⁻⁹, in the centre AATT part of the B-DNA dodecamer^{4,10-13}, and the fused pentagons within the TATA region of the A-DNA octamer⁵. These ordered structures are base-specific and are important elements in stabilizing regions of DNA with particular base sequences, and in influencing

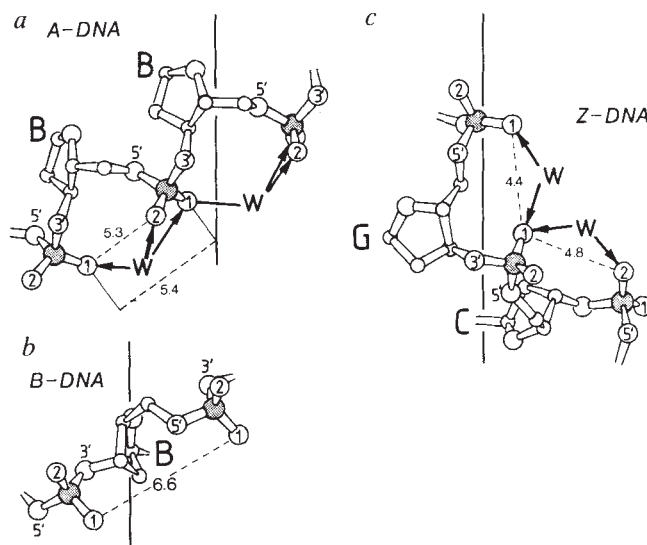


Fig. 1 Comparison of deoxyribose-3', 5'-diphosphate fragments in a, A-, b, B- and c, Z-DNA. Redrawn from corresponding figures in ref. 1. Vertical lines represent helical axes; phosphorus atoms indicated by shading, phosphate oxygens by numbering. Shortest distances between free phosphate oxygen atoms are given in Å units (see Table 1), W represents water molecules that bridge free phosphate oxygens in A- and Z-DNA. Capital letters B in A- and B-DNA and G, C in Z-DNA indicate C(1') sugar atoms where bases are attached.

Received 17 February; accepted 30 September 1986.

- Palmer, A. H. *J. biol. Chem.* **104**, 359-372 (1934).
- Pervais, S. & Brew, K. *Science* **228**, 335-337 (1985).
- Grosclaude, F., Mahe, M. F., Mercier, J. C., Bonnmaire, J. & Teissier, J. H. *Ann. Genet. Sel. anim.* **8**, 461-479 (1976).
- McKenzie, H. A. in *Milk Proteins: Chemistry and Molecular Biology* Vol. 2 (ed. McKenzie, H. A.) 257-330 (Academic, New York, 1971).
- Martial, J., Preaux, G. & Lontie, R. *Arch. int. Physiol. Biochim.* **79**, 842-843 (1971).
- Mainferme, F., Preaux, G. & Lontie, R. *Arch. int. Physiol. Biochim.* **79**, 840-841 (1971).
- McKenzie, H. A., Ralton, G. B. & Shaw, D. C. *Biochemistry* **11**, 4539-4547 (1972).
- Lyster, R. L. *J. Dairy Res.* **39**, 279-318 (1972).
- Jenness, R. in *Lactation: a Comprehensive Treatise*, Vol. 3, (eds Larson, B. L. & Smith, V. R.) 3-107 (Academic, New York, 1974).
- Jenness, R. *J. Dairy Res.* **46**, 197-210 (1979).
- Aschaffenburg, R. & Drewry, J. *Biochem. J.* **65**, 273-277 (1957).
- Miranda, G. & Pelissier, J. P. *J. Dairy Res.* **50**, 27-36 (1983).
- Tanford, C., Bunville, L. G. & Nozaki, Y. *J. Am. chem. Soc.* **81**, 4032-4036 (1959).
- Lovrien, R. & Anderson, W. F. *Arch. Biochem. Biophys.* **131**, 139-144 (1969).
- Futterman, S. & Heller, J. *J. biol. Chem.* **247**, 5168-5172 (1972).
- Stone, W. L. & Wishnia, A. *Bioinorg. Chem.* **8**, 517-529 (1979).
- Aschaffenburg, R., Green, D. W. & Simmons, R. M. *J. molec. Biol.* **13**, 194-210 (1965).
- Bolognesi, M., Liberatori, J., Oberti, R. & Ungaretti, L. *J. molec. Biol.* **131**, 411-413 (1979).
- Green, D. W. *et al. J. molec. Biol.* **131**, 375-397 (1979).
- Sawyer, L., Papiz, M. Z., North, A. C. T. & Eliopoulos, E. E. *Biochem. Soc. Trans.* **13**, 265-266 (1985).
- Rask, L. *et al. Scand. J. clin. Lab. Invest.* **40**, Suppl. 154, 45-61 (1980).
- Goodman, D. S. *The Retinoids* Vol. 2, 41-48 (1984).
- Newcomer, M. E. *et al. EMBO J.* **3**, 1451-1454 (1984).
- Jones, T. A. & Newcomer, M. E. *J. molec. Biol.* (submitted).
- Murthy, M. R. N., Reid, T. J., Siegnano, A., Tanaka, N. & Rossmann, M. G. *J. molec. Biol.* **152**, 465-499 (1981).
- Chothia, C. & Janin, J. *Biochemistry* **21**, 3955-3965 (1982).
- Godovac-Zimmermann, J., Conti, A., Liberatori, J. & Braunitzer, G. *Hoppe-Seyler Z. Phys. Chem.* **366**, 431-434 (1985).
- Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acid Res.* **12**, 387-395 (1984).
- Rao, S. T. & Rossmann, M. G. *J. molec. Biol.* **76**, 241-256 (1973).
- Rossmann, M. G. & Argos, P. *J. biol. Chem.* **250**, 7525-7537 (1975).
- Fugate, F. D. & Song, P.-S. *Biochim. biophys. Acta* **652**, 28-42 (1980).
- Horwitz, J. & Heller, J. *J. biol. Chem.* **249**, 4712-4719 (1974).
- Cogan, U., Kopelman, M., Mokady, S. & Shinitzky, M. *Eur. J. Biochem.* **65**, 71-78 (1976).
- Cambillau, C., Horjales, E. & Jones, T. A. *J. mol. Graph.* **2**, 53-54 (1984).
- Jones, T. A. *J. appl. Cryst.* **11**, 268-272 (1978).
- Eklund, H. *et al. J. molec. Biol.* **146**, 561-587 (1981).
- Hemley, R., Kohler, B. E. & Siviski, P. *Biophys. J.* **28**, 447-455 (1979).

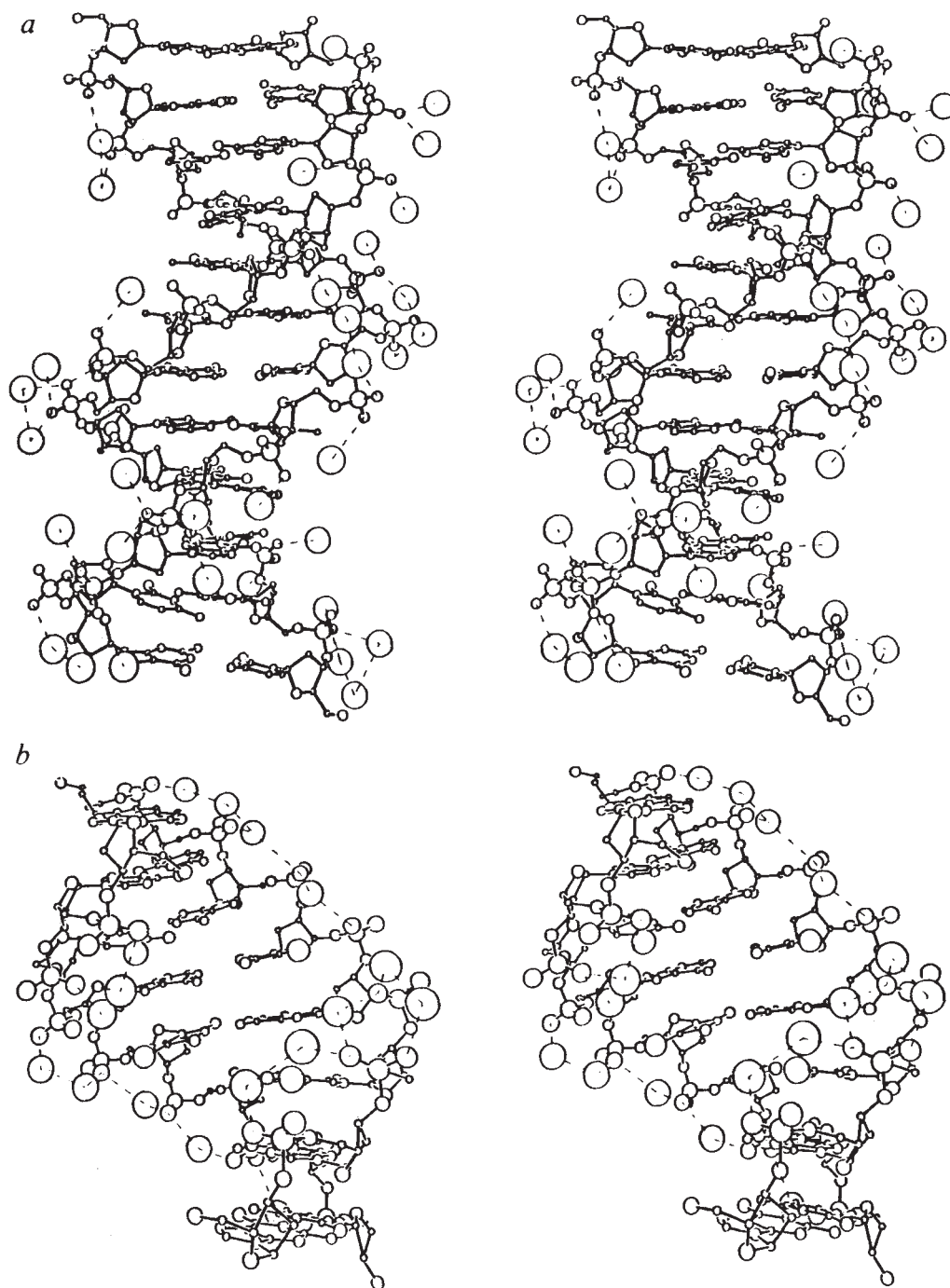


Fig. 2 *a, b*, DNA stereo-views. *a*, Stereoview showing the hydration of phosphate groups in the B-DNA dodecamer d(CGCGAATTAGCG) (ref. 10). Atoms are shown in order of size: water oxygen, P, O, N, C. Dashed lines, hydrogen bonds. Note the clustering of solvent molecules around phosphate groups and the lack of continuous water bridges. *b*, Hydration of phosphate groups in the A-DNA fragment d(GG^BUA^BUACC) (ref. 5). Of the 12 possible water bridges between free oxygen atoms of adjacent phosphate groups, 7 are actually formed within the cutoff limit of 3.5 Å for O-O separation in water. In one strand of the duplex, the water bridges are nearly continuous. The stereo-view is into the major groove.

conformational transitions of DNA with non-random sequence, such as the B → A transitions of DNA fibres with high A+T content¹⁴, and the B → Z transition¹⁵ of poly [d(G-C)]. However, it is not yet possible to extrapolate from these sequence-dependent ordered patterns to the more general mechanisms of conformational transitions in mixed-sequence DNA.

The most relevant difference between A-, B- and Z-DNA lies in the conformation of the sugar ring, which is related to the distance between the phosphate groups at the 3' and 5' ends of each nucleotide, and to the displacement of base pairs relative to the helical axis. The sugar pucker in B-DNA is 'soft' and extends beyond the O(4')-endo to C(1')-exo to C(2')-endo range or, in terms of pseudorotation phase angle¹⁶, from east to south, with the majority of sugars in C(1')-exo to C(2')-endo. These sugar puckers give rise to a wide phosphate-phosphate separation of 6.6 Å on average (Table 1 and Fig. 1). In contrast, in

A-DNA the sugar conformation is much more narrowly confined to C(3')-endo, and the distance between the phosphate groups is reduced to an average of about 6 Å. In Z-DNA, the sugar pucker alternates, with C(2')-endo for the pyrimidine nucleotides in *anti* conformation, and with O(4')-endo to C(1')-exo for the *syn* purine nucleotides. The resulting interphosphate distances differ considerably for the sequences listed in Table 1, but are comparable to A-DNA in the idealized Z-DNA, 5.9 Å for the pCp step and 6.0 Å for the pGp step.

Associated with the differences in phosphate-phosphate separation in A-, B-, Z-DNA are differences in distances between the successive free phosphate oxygens O(1) and O(2), that are further influenced by right- and left-handed helical twist, as shown in Table 1 and Fig. 1. On average, all interphosphate O-O distances exceed 6.6 Å in B-DNA. In A-DNA (omitting the very short sequence containing an iodinated cytidine residue d(¹CCGG)), they are considerably smaller; 5.4–5.7 Å for O(1)-

Table 1 Average intra-strand separations (Å) between free oxygen atoms O(1), O(2) in adjacent phosphate groups in A-, B-, Z-DNA

Type	Sequence	P-P	Intra-strand atomic distances (Å)				Resolution (Å)	Reference
			O(1)-O(1)	O(2)-O(2)	3'O(2)-O(1)5'	3'O(1)-O(2)5'		
A	d(GGGGCCCC)	6.0-6.6 6.2(0.8)	5.0-6.2* 5.7(0.9)	5.9-7.3 6.7(0.7)	7.0-7.9 7.6(0.8)	4.8-6.3* 5.5(0.8)	2.5	24
A	d(GG ^{Br} UA ^{Br} UACC)	5.6-6.3 6.0(0.7)	4.9-6.0* 5.5(0.7)	6.2-7.1 6.6(0.7)	7.0-7.9 7.4(0.8)	5.0-6.2* 5.5(0.8)	1.7	5
A	d(GGGGTCCC)	5.7-6.2 5.9(0.3)	5.0-6.0* 5.4(0.3)	6.3-6.7 6.5(0.1)	7.1-7.7 7.3(0.2)	4.9-5.8* 5.3(0.3)	2.2	25
A	d(GGGGCTCC)	5.5-6.6 5.8(0.5)	5.1-6.4* 5.6(0.5)	5.9-7.0 6.0(0.6)	6.8-7.9 7.5(0.4)	4.5-5.9* 5.4(0.5)	2.2	26
A	d(^l CCGG)	5.5-6.6 6.3(0.6)	5.9-7.3 6.6(0.6)	4.9-6.5 6.1(0.7)	7.0-8.1 7.6(0.6)	4.4-6.4 5.9(0.8)	2.2	27
B	d(CGCGAATTCGCG)	6.3-7.1 6.7(0.8)	5.9-7.5 6.8(0.8)	6.1-8.4 7.4(0.9)	6.7-8.9 8.1(1.0)	5.7-7.7 6.7(1.0)	2.3	11
B	d(CGCGAATTAGCG)	6.3-7.1 6.6(0.9)	5.8-7.8 6.7(0.9)	5.8-8.0 7.3(1.0)	7.1-9.0 8.0(1.0)	6.0-7.6 6.7(1.0)	2.5	10
B	d(CGCGAATTGCG)	6.0-7.2 6.6(0.3)	6.3-8.3 6.7(0.5)	6.1-8.7 6.9(0.6)	7.0-9.1 8.1(0.6)	5.8-7.3 6.6(0.6)	2.5	12
B	d(CGCAAATTCGCG)	5.9-7.1 6.6(0.5)	4.8-8.0 6.6(0.7)	5.9-9.0 7.4(0.7)	6.2-9.0 7.6(0.8)	5.8-8.3 6.7(0.6)	2.5	13
Z	d(^{Br} UGCGCG) step pCp	6.1-6.2 6.1(0.1)	7.0-7.3 7.2(0.1)	5.8-6.5 6.2(0.3)	5.0-5.2* 5.1(0.1)	8.1-8.2 8.1(0.1)	2.2	28
	step pGp	7.0-7.2 7.1(0.1)	5.3-5.8* 5.6(0.3)	8.8-9.6 9.2(0.4)	7.0-8.2 7.6(0.6)	7.2-7.6 7.5(0.4)		
Z	d(TGCGCG) step pCp	5.8-6.0 5.9(0.1)	6.5-7.0 6.8(0.3)	5.7-5.9 5.8(0.1)	4.4-5.0* 4.7(0.3)	7.7-7.8 7.8(0.1)	1.5	12
	step pGp	6.3-6.9 6.5(0.3)	4.6-6.2* 5.2(0.8)	8.5-9.0 8.7(0.3)	6.7-7.0 6.8(0.2)	6.7-7.4 7.0(0.4)		
Z	d(^{Br} CG ^{Br} CG ^{Br} CG) step pCp	5.5-5.9 5.8(0.2)	6.1-6.7 6.5(0.3)	4.7-6.0 5.5(0.6)	3.9-4.6* 4.3(0.3)	7.1-7.9 7.6(0.4)	1.5	7
	step pGp	6.1-6.5 6.3(0.2)	4.7-6.0* 5.2(0.6)	8.1-8.7 8.5(0.3)	6.7-7.3 7.1(0.2)	6.4-7.0 6.8(0.3)		
Z	d(CGCGCG) ideal step pCp step pGp	5.9 6.0	7.0 4.4*	5.5 8.2	4.8* 6.4	7.6 6.3		15

In A- and B-DNA, O(1) is that atom which is closer to the major groove side. Numbers give the spread of distances and the geometrical mean with standard deviation in parentheses. In the DNAs with base pair mismatches, the backbone geometry is virtually unchanged compared with all-Watson-Crick base pairs.

* Oxygen-oxygen contacts which are so close that they are most frequently (and almost systematically) bridged by water molecules. See also Fig. 1.

Table 2 Statistics of water binding to free phosphate oxygen atoms O(1), O(2) in different DNA fragments

Sequence*	I	II	III	IV	V	VI	VII
Helix type	A	A	B	B	B	Z	Z
Number of water molecules:							
Bound to O(1), O(2)	40	34	22	35	32	37	35
Forming bidentate intra-phosphate bridges	1	0	1	2	2	1	1
Forming monodentate interactions with O(1), O(2)	31	30	21	33	30	34	31
In one-water inter-phosphate bridges	7	4	0	0	0	5	3
In two-water inter-phosphate bridges	4	1	1	2	1	2	2
Inter-phosphate gaps bridged by one water (%)	58	33	0	0	0	63	38

The cutoff limit for O-O distances to water is 3.5 Å.

* I is d(GG^{Br}UA^{Br}UACC), ref. 5; II is d(GGGGCCCC), ref. 24; III is native d(CGCGAATTCGCG) ref. 11; IV is d(CGCGAATT^{Br}CGCG), MPD 7, ref. 4; V is d(CGCGAATTAGCG), ref. 10; VI is d(^{Br}CG^{Br}CG^{Br}CG), ref. 7; VII is d(^{Br}UGCGCG), ref. 28.

O(1), and 5.3 Å for 3'-O(1)-O(2)-5', compared to the other two distances O(2)-O(2) and 3'-O(2)-O(1)-5'. For Z-DNA in the idealized form, O(1)-O(1) distance for pGp (4.4 Å) and 3'-O(2)-O(1)-5' distance for pCp (4.8 Å) are short; O(2)-O(2) for pCp (5.5 Å) is intermediate; and all other O-O distances exceed 6.4 Å.

When the hydration schemes of the phosphate groups in the three different DNA conformations are compared, it is evident that in B-DNA the average oxygen-oxygen separation of about 6.6-8.1 Å is so large that each phosphate group can be surrounded by its own sphere of water molecules, and displays an independent hydration⁴. With a cutoff of 3.5 Å for O-O_{water} distances, one or two water molecules cannot bridge adjacent phosphates for the sequences given in Table 2. In contrast, in A- and Z-DNA, the oxygen atoms of adjacent phosphate groups

are so close together (Table 1) that they can be bridged by one or two water molecules⁵⁻⁷, as shown in Figs 1, 2 and in Table 2. The single water molecule bridges are always oriented in the same direction; in the 3' to 5' direction in A-DNA, the O(1) of the 3'-phosphate being linked by a water molecule to either the O(1) or O(2) oxygens of the adjacent 5'-phosphate, as these O-O distances are the shortest. In crystal structures the former is preferred because only in this case can the water molecule also interact with the functional groups of the bases in the major groove. For Z-DNA, the reverse direction of the bridges is found in the pCp step, O(2)-O(1) (Fig. 1). This directionality of water bridges is a result of the right- and left-handed twists of the respective DNA conformers, which determine the separations between phosphate oxygens.

B-DNA occurs under conditions of high water activity, and

A- and Z-DNA at low water activity; it appears that the hydration of A- and Z-DNA, where adjacent phosphate groups along the polyphosphate backbone are bridged by water molecules, is more economical and efficient than that of B-DNA, where each phosphate group is individually hydrated. As indicated by gravimetric¹⁷ and infrared spectroscopic¹⁸ studies and by crystal structure analyses of hydrated nucleotides¹⁹, the binding of water molecules to the charged phosphate oxygen atoms is energetically more favoured than binding to base and sugar atoms. Therefore, the phosphate hydration will have a substantial influence on the structure of DNA. If the water activity is high, the number of water molecules available for hydration of DNA is high, and the phosphate groups will be fully and individually hydrated. If salt or an organic polar solvent is added, water molecules are withdrawn from the DNA and hydration consequently becomes more economical. Economy is achieved by water molecules bridging adjacent phosphate groups, following a change in conformation of the sugar residues that results in the phosphate groups moving closer together. This transition between conformational states is cooperative, as is the formation of the bridges, which run along the entire sugar-phosphate backbone in A-DNA⁵. A comparable situation is seen in Z-DNA^{6,7} and we assume that this transition has the same basis as the B→A transition. Additionally, in Z-DNA the metal ions appear to play a significant role as in the crystal structures magnesium hexahydrate, cobalt hexammine, and sodium ions were found tightly coordinated with N(7) and O(6) of guanosine^{7,9,20}.

The concept of hydration economy finds further support if we consider the water content per base pair in the DNA double helices. It has been suggested, on the basis of gravimetric studies²¹ and theoretical considerations²², that an A·T base pair binds one more water molecule than a G·C base pair. This was confirmed in the crystal structure of the B-DNA dodecamer where water molecules are systematically trapped between the thymine methyl groups and one of the free oxygens of the 5' phosphates²³. Such an interaction can only occur if the nucleotide adopts a sugar conformation typical of B-DNA because then the base and the 5'-phosphate are farther apart than when the sugar is in the C(3')-endo conformation found in A-DNA. Therefore, DNA at high water activity adopting the B-type conformation is able to bind one additional water molecule at each thymidine, which confers additional stabilization. If the water activity is lowered, this (weakly) bound water molecule will be removed, and the transition from the B-type sugar puckering mode to the C(3')-endo form can take place. This again has to occur under cooperative control because otherwise the DNA would have both random B-DNA type and C(3')-endo pucker which would be unfavourable for the double helical structure.

Hence, it appears that the conformational transitions of mixed sequence DNA from B to A or B to Z are associated with the hydration of the phosphate groups which is less economical in B-DNA than in the A and Z forms. It is clear that other factors such as the spine of hydration in the minor groove of B-DNA and Z-DNA and the ordered water pentagons in the major groove of A-DNA also play a role, particularly in stabilizing specific base sequences.

O.K. is a member of the scientific staff of the Medical Research Council and the work was supported by an MRC programme grant. W.S. acknowledges a short-term EMBO fellowship to visit the University Chemical Laboratory, and support by Sonderforschungsbereich 9. We thank Dr E. Westhof for providing coordinates for Z-DNA and Dr R. E. Dickerson for constructive criticism of this paper.

4. Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. *J. molec. Biol.* **163**, 129-146 (1983).
5. Kennard, O. *et al. J. biomolec. struct. Dyn.* **3**, 205-225 (1986).
6. Westhof, E., Prangé, Th., Chevrier, B. & Moras, D. *Biochimie* **67**, 811-817 (1985).
7. Chevrier, B. *et al. J. molec. Biol.* **188**, 707-719 (1986).
8. Wang, A. H.-J., Hakoshima, T., van der Marel, G., van Boom, J. H. & Rich, A. *Cell* **37**, 321-331 (1984).
9. Quigley, G. J. in *Molecular Structure and Biological Activity* (eds Griffin, J. & Duax, W. L.) 317-332 (Elsevier, New York, 1981).
10. Brown, T., Hunter, W. N., Kneale, G. & Kennard, O. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2402-2406 (1986).
11. Dickerson, R. E. & Drew, H. R. *J. molec. Biol.* **149**, 761-786 (1981).
12. Kennard, O. *J. molec. struct. Dyn.* **3**, 205-225 (1985).
13. Hunter, W. N., Brown, T., Anand, N. N. & Kennard, O. *Nature* **320**, 552-555 (1986).
14. Arnott, S. & Selsing, E. *J. molec. Biol.* **88**, 509-521 (1974).
15. Wang, A. H.-J. *et al. Science* **211**, 171-176 (1981).
16. Altona, C. & Sundaralingam, M. *J. Am. chem. Soc.* **94**, 8205-8212 (1972).
17. Falk, M., Hartmann, K. A. Jr & Lord, R. C. *J. Am. chem. Soc.* **84**, 3843-3846 (1962).
18. Falk, M., Hartmann, K. A. Jr & Lord, R. C. *J. Am. chem. Soc.* **85**, 387-391 (1963).
19. Jeffrey, G. A., Maluszynska, M. & Mitra, J. *Int. J. biol. Macromol.* **7**, 336-347 (1985).
20. Gessner, R. V. *et al. Biochemistry* **24**, 237-240 (1985).
21. Tunis, M.-J. B. & Hearst, J. E. *Biopolymers* **6**, 1345-1353 (1968).
22. Goldblum, A., Perahia, D. & Pullman, A. *FEBS Lett.* **91**, 213-315 (1978).
23. Drew, H. R. & Dickerson, R. E. *J. molec. Biol.* **151**, 535-556 (1981).
24. McCall, M., Brown, T. & Kennard, O. *J. molec. Biol.* **183**, 385-396 (1985).
25. Kneale, G., Brown, T., Kennard, O. & Rabinovich, D. *J. molec. Biol.* **186**, 805-814 (1985).
26. Hunter, W. N., Kneale, G., Brown, T., Rabinovich, D. & Kennard, O. *J. molec. Biol.* (in the press).
27. Conner, B. N., Yoon, C., Dickerson, J. L. & Dickerson, R. E. *J. molec. Biol.* **174**, 663-695 (1984).
28. Brown, T., Kneale, G., Hunter, W. N. & Kennard, O., *Nucleic Acids Res.*, **14**, 1801-1809 (1986).

A protein on *Plasmodium falciparum*-infected erythrocytes functions as a transferrin receptor

Mario H. Rodriguez* & Michele Jungery†

Department of Tropical Public Health,
Harvard School of Public Health,
665 Huntington Avenue, Boston, Massachusetts 02115, USA

Several observations suggest that iron is essential for the development of malaria parasites¹⁻⁶ but there is evidence that the parasites in erythrocytes do not obtain iron from haemoglobin. The total haem level in parasitized erythrocytes does not vary during parasite development⁷, indicating that the iron-containing moiety of haemoglobin is not detectably metabolized. Although parasite proteases can degrade the protein part of haemoglobin in red cells^{8,9}, no parasite enzymes that degrade haemin have been identified. In mammalian cells, haemin is degraded to carbon monoxide and bilirubin by the enzyme haeme oxygenase¹⁰. This enzyme has not been found in malaria parasites¹¹. In fact haemin has been found to be toxic to parasite carbohydrate metabolism¹². Thus, iron apparently cannot be liberated from haemin and instead is sequestered in infected red cells as haemozoin, the characteristic pigment associated with malarial infection^{13,14}. If iron bound to transferrin is the source of ferric ions for malaria parasites within mature erythrocytes, then the parasite must synthesize its own transferrin receptor and localize it on the surface of the infected cell, because the receptors for transferrin are lost during erythrocyte maturation^{5,16}. Our results here suggest that *Plasmodium falciparum* synthesizes its own transferrin receptors enabling it to take up iron from transferrin by receptor-mediated endocytosis.

To determine whether human ferrotransferrin binds to parasitized cells, we conjugated this ferrotransferrin with fluorescein isothiocyanate (FITC) and observed the movement of the ligand from the cell surface to the parasitophorous vacuole inside erythrocytes for 30 minutes (Fig. 1). The transferrin was initially found in small vesicles in the cytoplasm of infected cells but within 30 min these fluorescent vesicles concentrated around the

Received 7 April; accepted 12 September 1986.

1. Saenger, W. in *Principles of Nucleic Acid Structure*, 220-297 (Springer, New York, 1983).
2. Dickerson, R. E. *et al. Science* **216**, 475-485 (1982).
3. Jurnak, F. A. & McPherson, A. (eds) *Biological Macromolecules and Assemblies Vol. 2: Nucleic Acids & Interactive Proteins* (Wiley, New York, 1985).

* Present address: Centro de Investigaciones de Paludismo, Tapachula, Chiapas, Mexico 30700.

† To whom correspondence should be addressed.

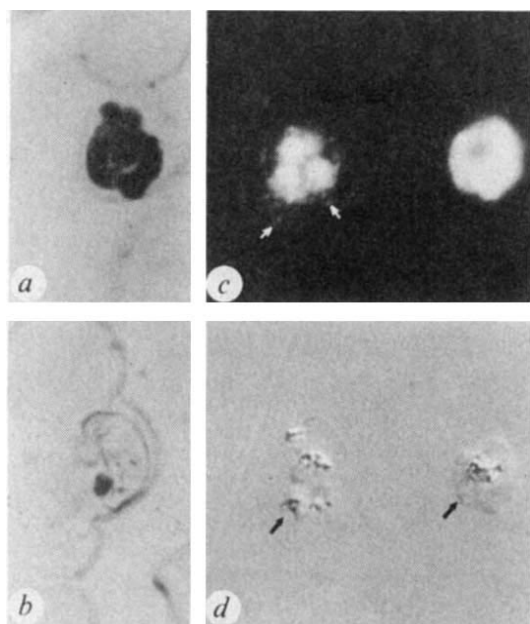


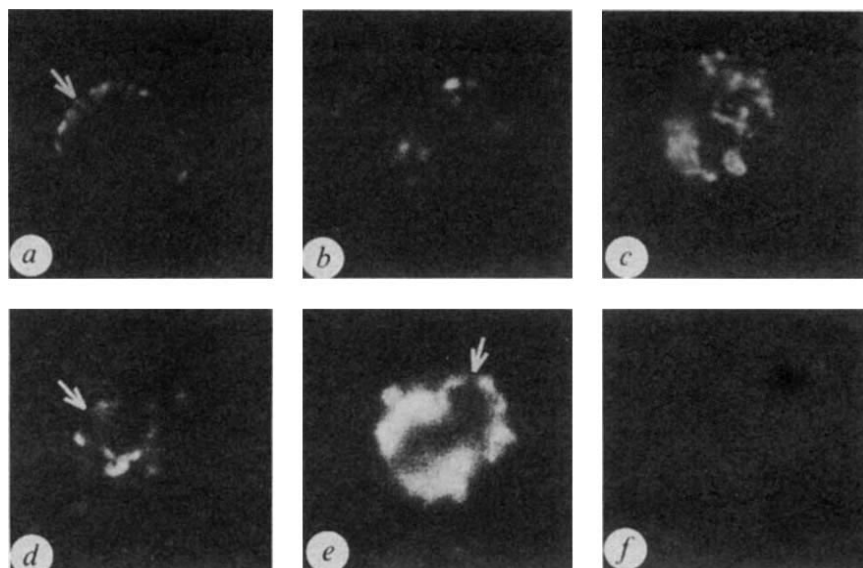
Fig. 1 Uptake of FITC-labelled ferrotransferrin by malaria parasites. Erythrocytes infected with *P. falciparum* (Palo Alto, Uganda strain) young trophozoites were obtained from cultures²⁶ by sedimentation on plasmagel²⁷. The infected cells were washed twice with RPMI 1640 (GIBCO) and 25 μ l of these cells were cultured in 100 μ l of complete culture medium (10% human serum, 0.4% glucose, 0.2% NaHCO_3 , 0.3% TES (*N*-Tris[hydroxymethyl]methyl-2-aminoethansulphonic acid), 0.004% gentamycin in RPMI 1640 medium) with $\sim 1 \text{ mg ml}^{-1}$ of FITC-ferrotransferrin. Transferrin (Calbiochem) was saturated with iron²⁸ and coupled to FITC (isomer I-Sigma) as described²⁹ with a final ratio FITC/protein of 2:1. Every 15 min cells were removed from culture, washed three times with cold RPMI and examined in a Zeiss ICM microscope with an epifluorescence condenser. Fluorescent transferrin was observed initially in small vesicles in the cytoplasm of the infected cells (*a*, arrow *b* and *c*). After 30 min at 37°C fluorescent material accumulated around the parasitophorous vacuole surrounding the parasite (*d* and *e*, arrows). Fluorescence was not evident on uninfected cells on the same slide. The specificity of FITC-transferrin uptake was further tested by showing competitive inhibition of uptake in the presence of an excess of unlabelled transferrin (a 1:100 dilution of the labelled transferrin): only a pale surface-fluorescence was observed (*f*).

parasitophorous vacuole surrounding the parasite in all the parasitized cells (Fig. 1*d, e*, arrows). Such fluorescence in parasitized cells was observable over a wide range of FITC-transferrin concentrations. When FITC-ferrotransferrin was diluted with an excess of unlabelled ferrotransferrin, little fluorescence was observed on the infected cells (Fig. 1*f*) indicating selective competition for this receptor. K562 cells, which carry transferrin receptors, were used to show the specificity of binding of FITC-transferrin. FITC-transferrin could be detected at concentrations from 1 mg ml^{-1} to less than $1 \mu\text{g ml}^{-1}$ except in the presence of unlabelled transferrin. Finally, iron-free apotransferrin, which has no affinity for transferrin receptor at neutral pH, did not bind to any of these cells.

We confirmed the presence of transferrin in vesicles and in or around the parasites in similar experiments using either peroxidase (Fig. 2*a, b*) or rhodamine (Fig. 2*c*) immunostaining of fixed parasite-infected cells treated with an anti-transferrin rabbit antibody. Most of the fluorescence was in or around the parasites or in vesicles (Fig. 2*c*, arrows) and not at the surface. These results agree with observations in mammalian cells that about 75% of the transferrin bound to receptors is found intracellularly¹⁷. Studies at the ultrastructural level to determine the precise location of the ligand and the receptor are in progress (D. Titus, M. Neutra, M. Jungery).

We used affinity chromatography and binding assays to identify the receptor(s) for transferrin in malaria-infected cells. Two different types of affinity chromatography were used to isolate the transferrin receptor. First, a ^{35}S -labelled parasite extract was passed through a column of anti-transferrin immunoglobulin coupled to Sepharose 4B. The SDS-polyacrylamide gel electrophoresis (PAGE) profile of parasite proteins from this column reveals a prominent band with relative molecular mass (M_r) of approximately 93,000 (93K), and a series of minor bands (Fig. 3, lane 2). In the second isolation procedure, ferrotransferrin was coupled to Sepharose 4B and washed in a way that releases the receptor from the transferrin/receptor complex in mammalian systems¹⁸. After incubation with the labelled parasite extract, the eluates contained few counts, indicating that most of the bound parasite material was not released. The major parasite protein that bound to the transferrin affinity column was again the 93K protein (Fig. 3, lane 4). The 93K protein did not bind to control columns coupled with either bovine serum albumin or ethanolamine. Therefore the most likely candidate for the parasite-derived transferrin receptor is the 93K transferrin-binding protein.

Fig. 2 *a, b*, Immunoperoxidase detection of transferrin on *P. falciparum*-infected erythrocytes; *c, d*, rhodamine-conjugated immunostains of *P. falciparum* infected cells. *P. falciparum* was cultured in complete medium containing 1 mg ml^{-1} ferrotransferrin. Infected cells were washed three times with RPMI and thin smears prepared. After fixation with 10% methanol in acetone, the slides were washed with 0.5% H_2O_2 to inactivate endogenous peroxidase³⁰, and treated with rabbit anti-transferrin serum and a peroxidase-labelled goat anti-rabbit IgG (Cappel). After development with aminoben-zidine- H_2O_2 , a red-brown stain was observed in infected erythrocytes around the parasitophorous vacuole (*a*). Specific anti-transferrin stain was seen on all infected cells. No stain was seen on non-infected cells or on infected cells treated with non-immune rabbit serum (*b*). *c, d*, Parasites were fixed and incubated with anti-transferrin antibody as in *a*. Slide 5 was washed in PBS with gelatin, then rhodamine-conjugated goat anti-rabbit serum was added. *c*, Anti-transferrin binding to vesicles (arrow) and accumulating around the parasite. *d*, A light-field (Nomarski optics) photograph of the same cells as in *c*. Only faint background fluorescence can be observed on normal erythrocytes seen in *c*.



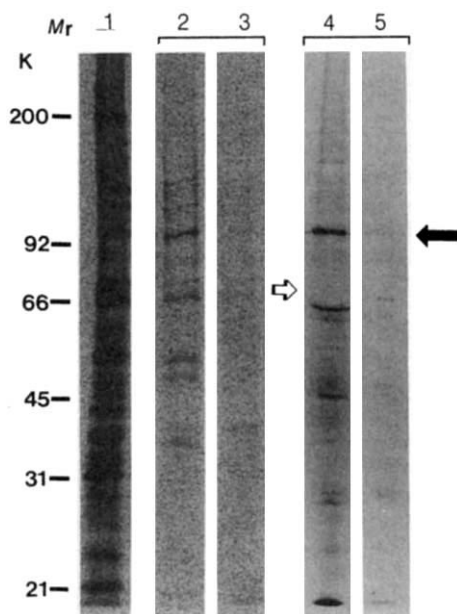


Fig. 3 Affinity chromatography of ^{35}S -methionine-labelled *P. falciparum* protein preparations on anti-transferrin-antibody bound to Sepharose or Fe-transferrin-Sepharose columns. *P. falciparum* trophozoites were metabolically labelled with ^{35}S -methionine for four hours. The parasites were harvested, solubilized in 0.15 M NaCl, 0.1 M PO_4 , pH 7.4 in phosphate-buffered saline (PBS) with 100,000 U ml^{-1} aprotinin, 1 mM iodoacetic acid, 5 mM TPCK, 5 mM TLCK and 2.0% Triton X-100. The preparation was incubated either with affinity purified rabbit anti-transferrin-antibody bound to Sepharose 4B or with Fe-transferrin-bound Sepharose 4B beads. Control columns included Sepharose 4B beads coupled with purified rabbit Ig (lane 3) or blocked with ethanolamine or bovine serum albumin (lane 5). Open arrow, location of the transferrin that was removed from the transferrin affinity Sepharose columns during boiling of the beads in SDS sample buffer.

Methods. The affinity column beads were incubated at room temperature for 3 h and then washed. The antitransferrin-Sepharose beads were washed with phosphate-buffered saline (PBS), 0.25% Triton X-100 until no counts were obtained from the washings. The Fe-transferrin beads were washed first with PBS, 0.5% Triton X-100, followed by 0.15 M Na acetate, 0.5% Triton X-100, pH 5, then 0.1 M Na citrate, 0.5% Triton X-100, 50 $\mu\text{g ml}^{-1}$ Desferrioxamine, pH 5, and finally by 0.1 M NH_4HCO_3 , 0.5% Triton X-200, 50 $\mu\text{g ml}^{-1}$ Desferrioxamine, pH 7.8. Beads from both columns were then boiled for 5 min in sample buffer, (0.1 M Tris-HCl, pH 7.5, 2% SDS, 10% mercaptoethanol, 2 mM ethylenediamine-tetraacetic acid (EDTA), 2 mM phenyl methyl sulphonyl fluoride (PMSF), 30% glycerol) and the supernatants analysed by SDS-PAGE³¹ and fluorography. Lane 1, the initial ^{35}S -methionine-labelled parasite preparation; lane 2, parasite proteins bound to and eluted from the anti-transferrin Ig-beads. The most prominent protein obtained has a molecular weight of 93K. Lane 4, parasite proteins that bind to the transferrin beads. The 93K protein is the most prominent protein binding to transferrin. Though it cannot be seen directly in this autoradiograph, the large amount of transferrin in lane 4 produced a displacement of other radiolabelled proteins in the 70K M_r (open arrow) region. The presence of transferrin was confirmed in Coomassie blue stain of this same gel. Lanes 3, 5 control affinity columns. Although there is some non-specific binding of parasite proteins the 93K protein is not present. Both affinity procedures identify a major protein band of M_r 93K (solid arrow) which binds to transferrin as the parasite transferrin receptor.

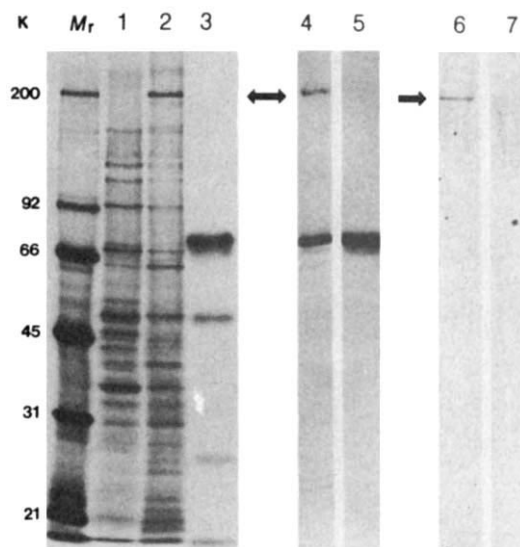


Fig. 4 Incorporation of ^{125}I -Fe-transferrin by *P. falciparum* in culture. Fluororadiograph of reduced SDS-PAGE of *P. falciparum* parasite preparations. *P. falciparum* trophozoite-infected erythrocytes were cultured in complete medium with 7 $\mu\text{Ci ml}^{-1}$ of ^{35}S -methionine (Amersham, 332 MBq ml^{-1}) for a total period of 4 h. After 3 h in culture, the preparation was divided into two aliquots, (samples 1, 2). A third was cultured for 3 h without ^{35}S -methionine. ^{125}I -labelled Fe transferrin was added to samples 2 and 3. The three cultures were incubated for a further hour at 37 °C. The parasites were harvested and washed three times with cold RPMI 1640. The parasitized red cells from samples 1 and 2 were lysed with 0.01 M Tris/HCl, pH 7.4 and spun at 10,000 r.p.m. for 30 min. This lyses the erythrocyte plasma membrane but leaves the intact parasites surrounded by the parasitophorous vacuole. The parasites were washed with 0.1 M Tris HCl pH 7.4 and extracts were prepared by adding solubilizing buffer (10 mM HEPES, 2.0% TRITON-X100, 2 mM PMSF, 5 mM L-1-tosylamide-2-phenylethyl chloromethyl ketone (TPCK), 5 mM *N*- α -p-tosyl-L-lysine chloromethyl ketone (TLCK), 50 $\mu\text{g ml}^{-1}$ leupeptin, 50 $\mu\text{g ml}^{-1}$ chymostatin). The extract was clarified by centrifugation at 100,000 g for 1 h. Supernatants containing extracted proteins were added to sample buffer (as described in Fig. 3 above) and run reduced conditions in SDS-PAGE gels. Lane 1, ^{35}S -methionine-labelled trophozoites from sample 1; lane 2, labelled trophozoites cultured in medium with ^{125}I -transferrin, sample 2; lane 3, ^{125}I -labelled Fe-transferrin. (Breakdown products were sometimes present in SDS gels.) ^{125}I -labelled ferrotransferrin alone was also added to sample 3, (young trophozoites in culture) for 1 h. Intact infected erythrocytes from sample 3 were harvested, washed, and dissolved in solubilizing buffer. The extract was clarified by centrifugation at 100,000 g for 1 h. The supernatant with the extracted proteins was added to sample buffer and resolved by SDS-PAGE. A 200K protein is present in both lanes 2 and 4, samples 2 and 3 to which ^{125}I -ferrotransferrin M_r 73K was added. The ^{125}I -ferrotransferrin (73K band, lane 4) is distinct from the ^{125}I -label which is present in a 200K protein band. We cannot at present explain how the 73K transferrin is associated with other molecules in SDS to give a 200K band. (Several procedures using different gel conditions are in progress.) Lane 5, the ^{125}I -labelled transferrin added to cells in lane 4. Arrows, ^{125}I -labelled F-transferrin-parasite protein complex (200K). The presence of transferrin in this 200K protein band was confirmed by immunoblotting the isolated parasite preparation from sample 2 (shown in lane 2) with purified rabbit anti-transferrin antibody and alkaline phosphatase-coupled goat anti-rabbit³². The result of this blot (lane 6) indicates that there was transferrin in the 200K protein. A control of non-infected mature erythrocytes cultured in medium with transferrin showed no reaction (lane 7).

We next determined the stage in the parasite's life-cycle where maximum uptake of ^{59}Fe bound to ^{125}I -labelled transferrin took place. By monitoring *P. falciparum*-infected cells from ring to schizont stages we found that the young trophozoite stage is most active, internalizing both radioactive labels which then accumulated in or around the parasite within the red cell. Further, after the erythrocyte membrane was removed from the infected cell by immune lysis, we found that both ^{59}Fe and ^{125}I -transferrin were still associated with the parasites isolated from the host cell.

In an effort to monitor changes in the presence of parasite proteins which might bind to transferrin, young trophozoites of *P. falciparum* were divided into two samples and metabolically labelled for four hours with ^{35}S -methionine. ^{125}I -transferrin was added at the end of three hours to sample 2. The parasites in samples 1 and 2 were collected at the end of four hours, then isolated from their red cells; this process was found to retain much of the parasite transferrin receptor. Solubilized parasite proteins from each sample were resolved by SDS-PAGE (Fig. 4). Lanes 1 and 2 show protein extracts from samples 1 and 2, (young trophozoites separated from the erythrocyte membranes); lane 3 shows the ^{125}I -transferrin used in sample 2. It was not possible to determine major differences in levels or the metabolically labelled protein between samples 1 and 2, but we did observe a 200K band in sample 2, suggesting that the ^{125}I -transferrin added to sample 2 was bound to a parasite protein, presumably the parasite transferrin receptor. This 200K transferrin complex is not present in quantities sufficient to be detectable in samples labelled only with ^{35}S -methionine but it must be different from the 195–210K protein seen in late stage *P. falciparum*-infected cells¹⁹ because that protein does not contain transferrin. To confirm the existence of a 200K band in sample 2, we incubated parasites (sample 3) for one hour with only ^{125}I -ferrotransferrin, after which the erythrocytes were collected intact. When the infected cells in sample 3 were solubilized and resolved by SDS-PAGE, we observed a ^{125}I -labelled 200K protein band distinct from the 73K transferrin band (Fig. 4, lane 4). Lane 5 shows the ^{125}I -transferrin which was added to sample 3. The 200K protein band was not seen when ^{125}I -ferrotransferrin was added to control cultures of uninfected mature erythrocytes.

The presence of transferrin in the 200K band was confirmed by immunoblotting the parasite-solubilized protein preparations with purified anti-transferrin immunoglobulin, then alkaline phosphatase conjugated anti-rabbit antibody (Fig. 4, lane 6). No reaction product was seen in the control normal erythrocytes solubilized and treated in the same way as infected cells (Fig. 4, lane 7). The control red cells were identical to those used in the parasite culture. We pretested these control red cells in several ways to ensure that they did not contain reticulocytes, and used the red cells (blood bank blood) in the *in vitro* culture of *P. falciparum* after two weeks storage at 4°C. We found that these cells did not take up FITC-transferrin or ^{125}I -transferrin, did

not bind anti-transferrin antibodies on Western blots or on fixed smears, and did not immunoprecipitate with antitransferrin antibody. Further, these red cells did not take up ^{35}S -methionine above background levels.

Mammalian transferrin receptors are composed of two disulphide-linked 93K proteins that, in turn, bind two molecules of transferrin²⁰. We do not yet know the nature of the parasite transferrin receptor/transferrin association. Based on the fact that a parasite derived transferrin receptor must bind mammalian transferrin and, consequently, may resemble the mammalian counterpart, we believe that the 93K protein synthesized by *P. falciparum* is the parasite's own transferrin receptor. It is, however, also possible that the 50K protein seen in the SDS-PAGE gels of affinity-purified parasite proteins is either a fragment of the 93K protein or perhaps functions together with this protein as a part of the receptor. Polyacrylamide gel electrophoresis of purified infected erythrocyte membranes, under nonreducing conditions, suggests such an association of these proteins (data not shown).

In each of the techniques used, we always found the 93K protein associated with or binding to human ferrotransferrin and not apotransferrin, strongly suggesting that this protein serves as a receptor for transferrin. We are now producing antibody to the 93K protein in mice following the procedure of Tung *et al.*²¹. This antibody will be used to determine whether the 93K protein, together with transferrin, is part of the 200K band. Preliminary Northern blot analysis of genomic *P. falciparum* DNA with the complementary (c)DNA clone of the human transferrin receptor indicates that the parasite has genes that code for this protein and that they are different from the human gene²³.

Several different approaches suggest that a *P. falciparum* parasite-derived molecule has transferrin receptor-like behaviour: (1) this molecule binds ferrotransferrin but not apotransferrin on parasitized erythrocytes; (2) FITC-transferrin uptake can be completely inhibited by an excess of unlabelled transferrin; (3) affinity chromatography with antitransferrin antibody or with transferrin reveals the presence of this specific parasite protein, and (4) the molecular weight of the complex of parasite receptor and transferrin from parasitized red cells is different from the transferrin/transferrin receptor complex described for reticulocytes^{15,16}. More importantly, our results indicate that malaria parasites transport proteins to the surface of the infected cells (see also refs 23, 24, 25), and use these proteins to obtain nutrients and other elements from the plasma. The parasite within the erythrocyte is in contact with macromolecules in its extracellular environment and is thus exposed to attack by immunological and pharmacological means.

We thank Drs H. Lodish, C. Mulford, D. Titus and J. Glass for their helpful comments and suggestions on this manuscript. This work was funded by the Rockefeller Foundation and by a grant from the Searle Scholars Program to M.J.

Received 18 February; accepted 28 October 1986.

- Masawe, A. E. J., Muindi, D. M. & Suai, G. B. R. *Lancet* **ii**, 314–317 (1974).
- Murray, M. S., Murray, A. B., Murray, M. B. & Murray, C. J. *Br. med. J.* **11**, 1113–1115 (1979).
- Raventos-Suarez, C., Pollack, S. & Nagel, R. L. *Am. J. trop. Med. Hyg.* **31**, 919–922 (1982).
- Harvey, P. W. J., Beil, R. G. & Nesheim, M. C. *Infect. Immun.* **50**, 932–934 (1985).
- Fritsch, G., Treumer, J., Spira, D. T. & Jung, A. *Exp. Parasitol.* **60**, 171–174 (1985).
- Pollack, S. & Fleming, J. *Br. J. Haematol.* **58**, 289–293 (1984).
- Ball, E. G., McKee, R. W., Anfinson, G. B., Cruz, W. O. & Geiman, Q. M. *J. biol. Chem.* **175**, 547–571 (1948).
- Moulder, J. W. & Evans, E. A. *J. biol. Chem.* **164**, 145–157 (1940).
- Levy, M. R., Siddiqui, W. A. & Chou, S. C. *Nature* **242**, 546–549 (1974).
- Landow, D., Callahan, Jr., E. W. & Schmid, R. *J. clin. Invest.* **49**, 914–925 (1970).
- Sherman, I. W. & Tanigoshi, L. *Biochemistry of Parasites* (ed. Slutsky, G. M.) 137–149 (Pergamon, Oxford, 1981).
- Keilin, D. & Hartree, E. F. *Biochem. J.* **41**, 403–406 (1947).
- Rudzińska, M. A. & Trager, W. *J. Biophys. biochem. Cytol.* **6**, 103–112 (1952).
- Deegan, T. & Macgregair, B. G. *Ann. trop. Med. Parasit.* **50**, 194–211 (1956).
- Van Bockmeier, F. M. & Morgan, E. H. *Biochim. biophys. Acta* **584**, 76–83 (1979).
- Lodish, H. F. & Small, B. J. *Cell Biol.* **65**, 51–64 (1975).
- Blail, J. D. & Bretscher, M. S. *EMBO J.* **1**, 351–355 (1982).
- Van der Lief, I. R., Stearne, P. A., Greco, B., Simpson, R. J. & Goding, J. M. *J. Immun.* **133**, 3220–3224 (1984).
- McBride, J. S., Newbold, C. I. & Anand, R. *J. exp. Med.* **161**, 160–180 (1985).
- Nunez, M. T., Glass, J. & Cole, E. S. *Biochem. biophys. Acta* **673**, 137–146 (1981).
- Tung, A. S., Shyr-te, S. & Nisonoff, A. *J. Immun.* **116**, 676–681 (1976).
- Rogers, J. & Jungery, M. (manuscript in preparation).
- Leech, J. H., Barnwell, J. W., Miller, L. H. & Howard, R. J. *J. exp. Med.* **159**, 1567–1575 (1984).
- Schmidt-Ullrich, R., Wallach, D. F. H. & Lightholder, R. J. *J. exp. Med.* **150**, 86–99 (1979).
- Wallach, D. F. H. & Conley, M. J. *molec. Med.* **2**, 119–136 (1977).
- Trager, W. & Jensen, J. B. *Science* **193**, 673–675 (1976).
- Reese, R. T., Langreth, S. G. & Trager, W. *Bull. WHO* **57** (Suppl.), 53–61 (1979).
- Bates, G. W. & Schlack, M. R. *J. biol. Chem.* **250**, 2177–2181 (1975).
- Markelonis, G. J. *et al. J. Cell Biol.* **100**, 8–17 (1985).
- Mason, D. Y., Abdulaziz, Z., Falini, B. & Stein, H. *Immunocytochemistry: Practical Applications in Pathology and Biology* (eds Polak, J. M. & Van Noorden, S.) 113–118 (Wright-PSG, Bristol, 1983).
- Laemmli, U. K. *Nature* **227**, 680 (1970).
- Burnette, W. N. *Analyt. Biochem.* **112**, 195–203 (1981).

Arc minute gravitational lenses and cosmic strings

PACZYŃSKI has suggested¹ that several large separation gravitational lens candidates could be explained by the existence of cosmic strings. These lens candidates are quasar pairs with separation between 1 and 4 arc min. Paczyński states that arc minute separation quasar pairs, which are lensed by long straight strings, are a prediction of cosmic string theory. Paczyński's work received a great deal of attention when preliminary evidence for a gravitational lens with a 2.6 arc min separation was published by Turner *et al.*² Because a lens with such a large separation is difficult to explain with unseen conventional objects and requires a large mass-to-light ratio for the lens, this lens candidate has been reported³ as evidence for a cosmic string. Subsequent observations suggest that the quasar pair is not lensed⁴. We point out here that such large separation lenses are not predicted by cosmic string theories.

The typical separation of the images of an object that is lensed by a cosmic string⁵⁻⁷ is $4\pi G\mu$ where $G\mu$ is the dimensionless mass parameter of the string. A gravitational lens with a separation of 2.6 arc min corresponds to $G\mu \approx 6 \times 10^{-5}$. While it is possible for a string configuration to produce gravitational lenses with image separations $>4\pi G\mu$ with favourable geometry, to increase the image separation by more than a factor of two or three above this value seems to require very unlikely conditions. While it was originally suggested⁸⁻¹⁰ that a value of $G\mu \approx 10^{-3}$ – 10^{-5} might be useful for galaxy formation, further work has shown that these values are too high, and probably already excluded by observation. Arc minute separation lenses are thus not a prediction of cosmic string theories. If some of these objects were lensed by long straight strings, it would create other cosmological problems, or suggest a basic misunderstanding of how cosmic strings should behave.

We now list the evidence against values of $G\mu > 10^{-5}$. Models where galaxies or clusters accrete around loops^{11,12} have obtained $G\mu \approx 2 \times 10^{-6}$. Thus the prediction of cosmic string models is that $G\mu \approx 10^{-6}$ not 10^{-5} . More serious problems for large values of $G\mu$ come from nucleosynthesis and anisotropy of the microwave background radiation. The most stringent limit on $G\mu$ comes from the requirement that the strings do produce so much gravitational radiation as to interfere with the successful primordial nucleosynthesis scenario¹³⁻¹⁵. This bound is stringent, $G\mu < 4 \times 10^{-6}$, but it can also be avoided by some types of cosmic strings that predominantly radiate massless goldstone

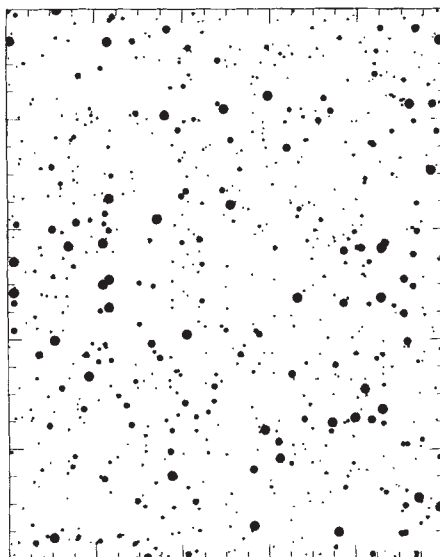


Fig. 1 Computer-generated image of 700 stars in a flat three-dimensional space, with a broad luminosity function.

bosons rather than gravitational waves¹⁶. However, these strings are not as attractive as seeds for galaxy formation, nor do they appear in any attractive particle physics models. A completely model-independent limit has been obtained by Kaiser and Stebbins¹⁷ who showed that $G\mu < 10^{-5}$ is required by the observed isotropy of the microwave background. Similar limits come from anisotropy due to string loops and the gravitational waves they produce^{18,19}. While estimating the errors on these theoretical limits is difficult, we feel that at least the limit of Kaiser and Stebbins is actually conservative. As they stated, the limit given is for a "minimal model" which, for example, does not include the increase in the r.m.s. temperature fluctuation due to the superposition of strings which will occur in any model. Thus, the limit $G\mu < 10^{-5}$ is quite firm, so the predicted separation of images lensed by a cosmic string is definitely <1 arc min.

This work was supported by the Department of Energy, contract DE AC03 76SF00515.

DAVID P. BENNETT

MS 209,
Fermilab,
P.O. Box, 500, Batavia,
Illinois 60510, USA

ALBERT STEBBINS

Physics Department,
University of California, Berkeley,
California 94720, USA

5. Vilenkin, A. *Astrophys. J.* **282**, L51–53 (1984).
6. Gott, J. R. *Astrophys. J.* **288**, 422–427 (1985).
7. Frieman, J. in *Inner Space Outer Space* (eds Kolb, E. W., Turner, M. S., Lindley, D., Olive, K. & Seckel, D.) 540–542 (University of Chicago Press, 1986).
8. Zel'dovich, Y. B. *Mon. Not. R. Astr. Soc.* **192**, 663–667 (1980).
9. Vilenkin, A. & Shafi, Q. *Phys. Rev. Lett.* **51**, 1716–1719 (1983).
10. Turok, N. *Nucl. Phys. B* **242**, 520–541 (1984).
11. Turok, N. & Brandenberger, R. *Phys. Rev. D* **33**, 2175–2181 (1986).
12. Stebbins, A. *Astrophys. J.* **303**, L21–25 (1986).
13. Davis, R. L. *Phys. Lett.* **161B**, 285–288 (1985).
14. Bennett, D. P. *Phys. Rev. D* **33**, 872–888 (1986) *erratum* (in the press).
15. Bennett, D. P. *Phys. Rev.* (in the press).
16. Davis, R. L. *Phys. Rev. D* **32**, 3172–3177 (1985).
17. Kaiser, N. & Stebbins, A. *Nature* **310**, 391–393 (1984).
18. Brandenberger, R. H. & Turok, N. *Phys. Rev. D* **33**, 2182–2195 (1986).
19. Traschen, J., Turok, N. & Brandenberger, R. *Microwave Anisotropies from Cosmic Strings*, ITP Preprint, NSF-ITP-85-110 (1985).

PACZYŃSKI REPLIES—In my letter¹ I intended to emphasize two issues. Firstly, it is unlikely that cosmic strings could be discovered by finding pairs of images. I pointed out that there is a more striking effect of a string lensing: the images of some galaxies would appear as if they were cut with a sharp edge. Secondly, I pointed out that with the images separated by a few arc minutes the expected difference in the light travel time between the two images may be a few hundred or even a few thousand years. We do not know how stable are the broad line spectra of quasars, and I presented some observational reasons why they may, in fact, be variable. Therefore, I proposed that the forbidden lines in the spectra of pairs of quasars may help to decide between a cluster of quasars and a gravitational lens hypothesis. I quoted some papers (not mine) that suggested that strings may produce pairs of images separated by up to a few arc minutes, but this was just one of the many possibilities I listed. Therefore, it is surprising that Bennett and Stebbins give me so much credit for the cosmic string as the cause of the 2.6 arc min lens candidate.

I would like to take this opportunity to re-emphasize the same two points. Figure 1 shows a computer-generated field with about 700 'stars' randomly placed in a flat three-dimensional space, with a broad luminosity function. Half-way between us and the 'horizon' there is a straight 'cosmic string' that produced 27 pairs of images. Readers are encouraged to find the pairs and the string. This can be done, but the presence of the 'string' is not striking. That was my first point¹.

I do not think that it is possible to prove or disprove the gravitational lens hypothesis for the 2.6 arc min candidate of Turner *et al.* by analysing their spectra, as the forbidden lines are too weak to be seen. Fortunately, the quasars (or the quasar) are relatively close, with redshift, $z=1.01$. With present technology it is

1. Paczyński, P. *Nature* **319**, 567–568 (1986).
2. Turner, E. L. *et al.* *Nature* **321**, 142–144 (1986).
3. Gott, J. R. *Nature* **321**, 420–421 (1986).
4. Shaver, P. A. & Christiani, S. *Nature* **321**, 585–586 (1986).

possible to see, although with some difficulty, galaxies at this redshift and beyond. Galaxies have two obvious advantages over quasars for our purpose: their space density is much higher, and they do not vary on a timescale shorter than a few billion years. Therefore, if there is a giant gravitational lens then we should be able to see many multiply-lensed galaxies. If these are discovered we shall know there is a lens, and we shall be able to study the lens structure. If the multiply-lensed galaxies are not found then we shall have to conclude that there is no lens. The task is obviously difficult, and it may take years to complete, but I think that any proof may come only from studies of far away galaxies in the field of the 2.6 arc min lens candidate.

P. PACZYŃSKI

*Princeton University Observatory,
Peyton Hall, Princeton,
New Jersey 08544, USA*

1. Paczyński, P. *Nature* 319, 567-568 (1986).

'Capture' of stereopsis by illusory contours

RAMACHANDRAN and Cavanagh¹ reported an interesting effect in which disparate subjective contours capture regions of zero disparity enclosed within these contours but not regions that lie outside the contours. They suggest that this phenomenon challenges the current computational models of stereo matching. However, the effect is not as general as their article implies: it occurs only in the presence of ambiguity posed by periodic structures. When non-periodic backgrounds are used the stereo capture does not occur (Fig. 1 bottom). This suggests that the phenomenon may have something to do with the treatment of periodic structures by the visual system².

Most current computational models of stereo matching³⁻⁵ are based on point-by-point matching of features and use a set of assumptions or constraints (based on some very general properties of the environment) to ease the computational burden associated with solving the often massive local ambiguity problem. Surfaces containing periodic texture cannot be successfully matched locally. For example, a periodic texture on a frontoparallel plane will produce a set of peaks in the cross-correlation surface, each corresponding to a possible match. Locally, there is not enough information to decide which of these possibilities is the correct one. True correspondences can be obtained only by using more global information or by incorporating information originating outside the stereomatching system. This is true for any matching mechanism relying on local computations.

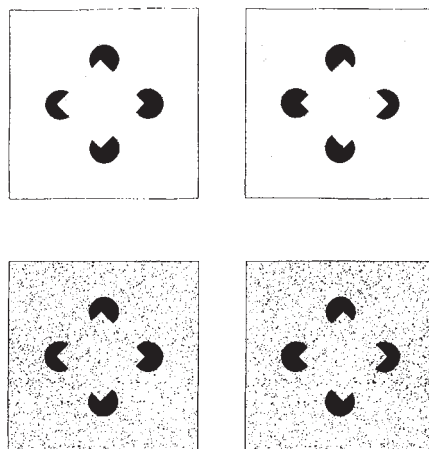


Fig. 1 Depth carried by disparate subjective contours is attributed to periodic (top) but not to non-periodic (bottom) texture elements enclosed by these contours. The stereograms were constructed for viewing with free fusion by crossing the eyes in front of the displays.

The problem of multiple 'local maxima' has been often attacked by invoking multi-resolution strategies in which low spatial frequency matches are used to disambiguate the high frequency ones⁴. However, such strategy is not of much use in the case of extended periodic structures. Another, more general, way to solve this problem is to use other cues, for example, accommodation⁶, to bias disparity. The idea of a set of depth maps cooperating to converge on a state in which all maps agree and present sharp and accurate three-dimensional localization of features is a powerful notion, the potential of which has yet to be fully explored in the machine vision research.

Current computational theories of stereopsis do not address the qualitative or quantitative aspects of binocular depth perception. Rather, they attempt to understand only its first step: the detection of binocular disparities. Not all of them assume that binocular disparities are directly linked to our perception. That is to say, not all of them assume that our perception of depth from stereograms are isomorphic to the state of the neural disparity network. It is possible, to use an analogy, that binocular disparities are to the stereoscopic depth perception what the

output of retinal receptors is to the perception of brightness.

Rather than challenging current computational models of stereomatching, stereoscopic capture illuminates, I believe, mechanisms underlying more cognitive or configurational aspects of stereoscopic processing.

1. Ramachandran, V. S. & Cavanagh, P. *Nature* 317, 527-530 (1985).
2. Mitchison, G. J. & McKee, S. P. *Nature* 315, 402-404 (1985).
3. Mayhew, J. E. W. & Frisby, J. P. *Perception* 9, 69-87 (1980).
4. Marr, D. & Poggio, T. *Proc. R. Soc. B204*, 301-328 (1979).
5. Prazdny, K. *Biol. Cybern.* 52, 93-99 (1985).
6. House, D. in *Proc. Visuomotor Coordination in Frog and Toad Workshop* (ed. Arbib, M.) Tech. Rep. 82-16 (University of Massachusetts, Amherst, 1982).

K. PRAZDNY

*Machine Perception Group,
Artificial Intelligence Laboratory,
FMC Corporation,
1185 Coleman Avenue,
Box 580, Santa Clara,
California 95052, USA*

RAMACHANDRAN REPLIES—in our article on stereoscopic capture¹ we argued that existing stereo-algorithms could not account for our findings although it is possible that some of them could be modified to do so. This is quite consistent with Prazdny's interesting suggestion. Since disparity of illusory contours can influence the matching of finer elements in a stereogram (Fig. 1, top), we concluded that monocular contours are involved in stereopsis^{1,2}. Contrary to Prazdny's suggestion, however, we believe this to be a very general effect and 'capture' happens to be just one way of demonstrating their role. Prazdny also suggests that the effect is specific to 'periodic' patterns, a possibility that we ourselves consider (page 528) when discussing the phase sensitivity of stereo-capture. However, it is almost certainly the lack of ambiguity (not lack of periodicity) that prevents capture in Prazdny's Fig. 1 (lower).

Furthermore, there are also three other aspects of our effect that are not easily handled by existing computational models. First, stereo capture can be seen even in rivalrous displays^{1,2} (especially when flashed briefly) and not just in peri-

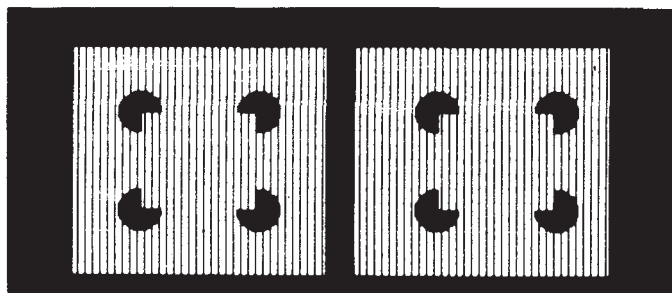


Fig. 2 Stereo capture is not seen if disparity is introduced between the disks themselves.

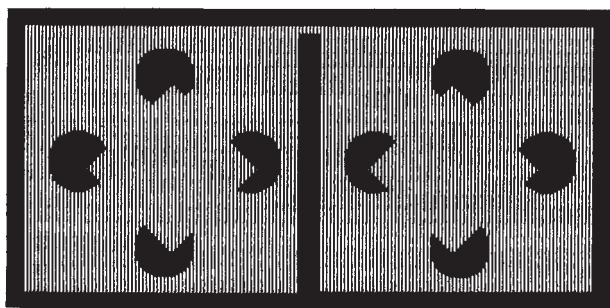


Fig. 3 Occlusion must be an inherent part of the stereoscopic matching process. The four black disks now become 'holes' through which one can see the four deep corners of a partially hidden diamond and the corners pull the corresponding lines of wallpaper with them.

odic displays. Second, Prazdny suggests that disparity signals from the 'unambiguously matched borders' help obtain true correspondences. This conjecture cannot be valid, at least in its simple form, as Fig. 2 does not yield stereo capture. (Here the disparity is between the disks themselves and not just between the cut sectors.) Hence construction of a three-dimensional surface and the presence of a steep disparity gradient are prerequisites; the mere propagation of depth signals will not suffice. Third, we find that reversing the sign of disparity does not simply reverse perceived depth. When the pictures of the two eyes are interchanged to convey uncrossed disparities (Fig. 3), a qualitatively different percept emerges. The black disks now look like holes through which one can see the four corners of a partially occluded square (diamond) and these deeper corners pull the corresponding lines with them. The illusory contours are now seen to complete the holes rather than the square. Here, the illusory contours are created after stereo matching and must, in their turn, influence the subsequent stereo matching of finer elements.

Effects similar to stereo capture can also be observed in apparent motion (motion capture)³ but the effect is not phase sensitive and has a much higher tolerance for uncorrelated textures.

Any visual process can in principle be simulated in a computer but our results suggest that the brain solves the correspondence problem (in motion and stereopsis) by using a set of short cuts⁴ rather than elaborate computation of the kind implied by artificial intelligence researchers. During phylogeny the visual system may have acquired many such tricks which were adopted not for their aesthetic appeal, versatility or mathematical elegance but simply because they worked—an idea that we call the utilitarian theory of perception⁵. Artificial intelligence can be useful in rigorously reformulating^{6,7} perceptual problems, but the mechanisms used by the brain to solve these problems are best revealed by old-fashioned psychophysics and neuro-

anatomy^{5,8}. Capture probably results from the fact that the motion (or depth) of certain salient image features is first extracted by the magnocellular pathways and then attributed to finer features that excite the parvocellular pathways. The nature of this interaction merits further attention.

V. S. RAMACHANDRAN

*Department of Psychology C-009,
University of California at San Diego,
La Jolla, California 92093, USA.*

1. Ramachandran, V. S. & Cavanagh, P. *Nature* **317**, 527-530 (1985).
2. Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. *Nature* **242**, 412-414 (1973).
3. Ramachandran, V. S. & Cavanagh, P. *Vis. Res.* (in the press).
4. Ramachandran, V. S. & Anstis, S. M. *Scient. Am.* **254**, 102-110 (1986).
5. Ramachandran, V. S. *Perception* **14**, 97-103 (1985).
6. Marr, D. *Vision* (Freeman, San Francisco, 1982).
7. Ullman, S. *The Interpretation of Visual Motion* (MIT, Cambridge, Massachusetts, 1979).
8. Livingstone, M. S. & Hubel, D. H. *J. Neurosci.* **4**, 309-356 (1984).

Is the Gulf of Tadjoura a platelet boundary or complex rift zone?

CHOUKROUNE *et al.*¹ have provided new data for the structure of the Gulf of Tadjoura (central Afar), which supposedly separates the Danakil and Somali platelets. Their observations (orientation and activity of a NW-SE fault set) made during dives with the *Cyana* submersible, are at odds with the model of orthodox plate-tectonics boundary, which had been developed in the central Afar and in the Gulf, and was previously widely accepted. Choukroune *et al.* question the westwards opening of the Gulf by rifting: their new interpretation of the Arabia and Africa plate-tectonics relationship does not imply that the Afar triangle is a multi-platelet building, but that it could represent a diffuse relay zone between the Aden Rift and the Red Sea Rift.

The new data support observations (unpublished) made during earlier cruises by Boucarut *et al.* Their complete tectono-bathymetric chart of the Gulf of Tadjoura

shows that the Maskali-Obock trough and the Tadjoura trough are located on either side of a subcontinuous tectonic lineament (major fault zone), running ENE-WSW (N 60°) from 43°20' to 43°. Both troughs are delineated either by ENE-WSW, NW-SE, or EW fault zones, which form some important (Recent) submarine cliffs. Moreover, the presence of active (NW trending) extensional fault zones, some of them crossing the Gulf and cropping out on both sides², and of faulted Middle Pleistocene sediments in the Obock trough and on the northern side of the Gulf³ had been mentioned.

Continuation of the major structures on land into the Gulf suggests that the Gulf floor is composed mainly of broken continental crust, separated by relatively narrow and irregular shaped rift zones filled with young volcanics².

The various orientations of the fault sets in the Gulf should be compared with those of the nearby land area, whose successive motions and combinations with volcanic occurrences during Plio-Pleistocene times are well known⁴. On land, the structural dominance of the 'Aden type' fractures (N 80°-100°) diminishes westwards: likewise, in the Gulf, it is taken over by that of the NW-SE and ENE-WSW faults. This explains the zigzag shape for the (supposed) boundary.

So, even if the platelet sketch for the central Afar has to be reviewed, various data indicate that the westwards tearing interpretation of the Gulf of Tadjours⁵⁻⁷ should not be eliminated.

M. BOUCARUT
M. CLIN

*University of Bordeaux 3,
33405 Talence Cedex, France*

1. Choukroune, P. *et al. Nature* **319**, 396-399 (1986).
2. Boucarut, M., Clin, M. & Dars, R. *C.r. hebdom. Séanc. Acad. Sci. Paris* **285D**, 491-494 (1977).
3. Clin, M. & Pelissier-Hermitte, G. *C.r. Campagne Orgon IV, CEPM-CNEXO*, 183-188 (CNRS, Paris 1979).
4. Boucarut, M., Clin, M., Pouchan, P. & Thibault, C. *Geol. Rdsh.* **74**, 123-137 (1985).
5. Boucarut, M. & Clin, M. *C.r. hebdom. Séanc. Acad. Sci. Paris* **282D**, 1145-1148 (1976).
6. Courtillot, V. *Phys. Earth planet. Inter.* **21**, 343-350 (1980).
7. Boucarut, M. & Clin, M. *Atti Convegno Lincei, Roma* **47**, 541-554 (1980).

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.